Technical University of Košice

Faculty of Electrical Engineering and Informatics

SCYR

23rd Scientific Conference of Young Researchers
Proceedings from Conference

ISBN 978-80-553-4377-8

2023

Sponsors & Organizers

SCYR 2023: 23rd Scientific Conference of Young Researchers
Proceedings from Conference

Published: Faculty of Electrical Engineering and Informatics
Technical University of Košice
Edition I, 196 pages, number of CD Proceedings: 50 pieces

Editors: Prof. Ing. Alena Pietriková, CSc.
Assoc. Prof. Ing. Emília Pietriková, PhD.

ISBN 978-80-553-4377-8

Scientific Committee of SCYR 2023

General chair: Prof. Ing. Liberios Vokorokos, PhD.

Editorial board chairman: Prof. Ing. Alena Pietriková, CSc.

Committee Members & Reviewers:

Prof. Ing. Roman Cimbala, PhD.
Prof. RNDr. Vladimír Lisý, DrSc.
Prof. Ing. Ján Paralič, PhD.
Prof. Ing. Daniela Perduková, PhD.
Prof. Ing. Alena Pietriková, CSc.
Assoc. Prof. Ing. František Babič, PhD.
Assoc. Prof. Ing. Jaroslav Džmura, PhD.
Assoc. Prof. Ing. Ján Genči, PhD.
Assoc. Prof. Ing. Karol Kyslan, PhD.
Assoc. Prof. Ing. Ján Papaj, PhD.
Assoc. Prof. Ing. Emília Pietriková, PhD.
Ing. Juraj Bujňák, PhD., Siemens Healthineers Košice

Organizing Committee of SCYR 2023

Members: Prof. Ing. Alena Pietriková, CSc.
Ing. Ivana Olšiaková
Assoc. Prof. Ing. Emília Pietriková, PhD.

Contact address: Faculty of Electrical Engineering and Informatics
Technical University of Košice
Letná 9
040 01 Košice
Slovak Republic

Foreword

Dear Colleagues,

SCYR (Scientific Conference of Young Researchers) is a scientific event focused on exchange of information among young researchers from Faculty of Electrical Engineering and Informatics at the Technical University of Košice – series of annual events that was founded in 2000. Since 2000, the conference has been hosted by FEEI TUKE with rising technical level and unique multicultural atmosphere. The 23rd Scientific Conference of Young Researchers (SCYR 2023) was held on April 20, 2023 at University Conference Centre, Technical University of Košice. The mission of the conference, to provide a forum for dissemination of information and scientific results relating to research and development activities at the Faculty of Electrical Engineering and Informatics, has been achieved. Approx. 70 participants, mostly by doctoral categories, were active in the conference.

Faculty of Electrical Engineering and Informatics has a long tradition of students participating in skilled labor where they have to apply their theoretical knowledge. SCYR is an opportunity for doctoral and graduating students to train their scientific knowledge exchange. Nevertheless, the original goal is still to represent a forum for the exchange of information between young scientists from academic communities on topics related to their experimental and theoretical works in the very wide spread field of a wide spectrum of scientific disciplines like informatics sciences and computer networks, cybernetics and intelligent systems, electrical and electric power engineering and electronics.

Traditionally, contributions can be divided in 2 categories:

- Electrical & Electronics Engineering
- Computer Science

with approx. 70 technical papers dealing with research results obtained mainly in the University environment. This day was filled with a lot of interesting scientific discussions among the junior researchers and graduate students, and the representatives of the Faculty of Electrical Engineering and Informatics. This Scientific Network included various research problems and education, communication between young scientists and students, between students and professors. Conference was also a platform for student exchange and a potential starting point for scientific cooperation. The results presented in papers demonstrated that the investigations being conducted by young scientists are making a valuable contribution to the fulfillment of the tasks set for science and technology at the Faculty of Electrical Engineering and Informatics at the Technical University of Košice.

We want to thank all participants for contributing to these proceedings with their high quality manuscripts. We hope that conference constitutes a platform for a continual dialogue among young scientists.

It is our pleasure and honor to express our gratitude to our sponsors and to all friends, colleagues and committee members who contributed with their ideas, discussions, and sedulous hard work to the success of this event. We also want to thank our session chairs for their cooperation and dedication throughout the entire conference.

Finally, we want to thank all the attendees of the conference for fruitful discussions and a pleasant stay in our event.

Liberios VOKOROKOS
Dean of FEEI TUKE

April 20, 2023, Košice

Contents

Design of protection system for microgrids

[1]*Róbert Štefko (3rd year)*
Supervisor: [2]Michal Kolcun

[1,2]Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic

[1]robert.stefko@tuke.sk, [2]michal.kolcun@tuke.sk

***Abstract*—The aim of this paper is to compare the possibility of using the protection system currently in use with a new centralized system for microgrid protection. The research indicates that the current protection system cannot be used in microgrids and will be replaced by a new Phasor Measurement Unit (PMU) and centralized data processing and state evaluation.**

***Keywords*—protection system, communication of IEDs, microgrids, protection relay.**

I. Introduction

The continuous deployment of renewable electricity sources (RES) into the distribution grid is starting to have an impact on the current protection, control, and fault location systems. This change is shifting the conventional mode of generation from large power plants to very small local power plants. In the current system, we use only one direction, from generation to consumption, but in the smart grid, this changes to a two-way flow of energy, with consumers themselves participating in generation. For this reason, the size of the microgrid will play a significant role, and so will the deployment and number of resources in it. Such deployment of RES changes the direction of power flow, voltage, and frequency ratios. Research suggests that microgrids are the solution to the RES problem. The main issue with microgrids is the selection and placement of local power sources. A major challenge for researchers dealing with this area is the design of a new protection and control system.

The main task of the protection system is to ensure that the protected part or equipment is not subjected to abnormal conditions, such as overload or short circuit, which would significantly shorten the lifetime of the equipment or protected part. For this reason, it is necessary to know the relationship of the equipment or section to the environment. The same principle must be applied to the microgrid [1].

My motivation for tackling the research task was to contribute the knowledge gained to the advancement of research in the field of protection systems.

II. Overview of the Research Outputs

Analyzing the current state of power plants on the territory of the Slovak Republic and assessing the share of installed renewable electricity sources for use in microgrids were of the first research objectives [2]. Subsequently, a survey was conducted on the applicability of microgrid power source mixes in projects worldwide [3]. The possibility of using existing technologies in practice and their further use in microgrids also needed to be analyzed [4]. One of the first devices to be examined was the line fault location sensor. These work on different principles [5]. Another significant development was to create models in Simulink of renewable energy sources that could be used in Slovenia, such as photovoltaic power plants, battery stations, hydroelectric power plants, and biogas power plants [2], [6]. The model was also created to test the possibility of the use of intelligent electronic devices, specifically for overcurrent protection with mutual communication, directional overcurrent protection with mutual communication, and differential protection [7].

III. Functionality Test of the Protection System

Simulink was used as the test environment, and models of the individual renewable energy sources were also created for this purpose. The test topology of the designed microgrid consisted of three ring networks. This minimized the disconnection of customers supplied by each power plant in the event of a fault. In the developed microgrid topology, eleven faults (F1-F11) were considered. They were located on each interconnection line as shown in Fig. 1.

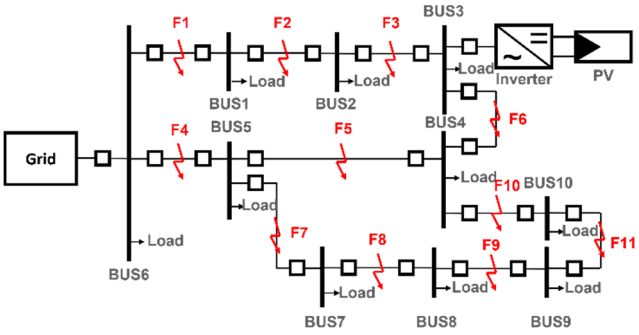


Fig. 1. Tested microgrid topology.

The individual loads that were placed at each of the stations represented the dynamic load for which a realistic load model had been designed. These dynamic loads have two dynamic characteristics, which have been distributed in a random manner among the individual electrical stations.

A. Differential protection relay

Differential protection was the first consideration for individual substation protection and line protection was the first proposal for its use in microgrids. The schematic diagram of the differential protection model for use in the protection of the line is shown in Fig. 2. For substation protection, it is supplemented by additional current inputs.

The results of the simulations confirm the possibility of using the system in practice, due to the fault-free operation for the eleven fault conditions that were tested for different areas.

The only disadvantage of using differential protection for line protection is the somewhat complicated wiring of a remotely located, instrumented current transformer which is located on the other side of the line, several tens of kilometers away.

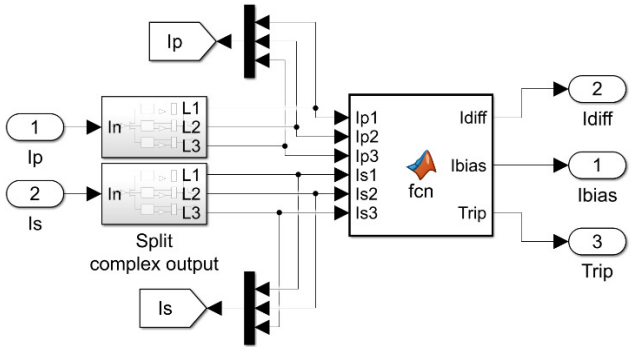


Fig. 2. Model diagram of differential protection relay [7].

B. Overcurrent protection relay

Overcurrent protection represents the second option considered for protecting the line and is also currently used for protecting the line. The schematic diagram of the overcurrent protection model used for line protection is shown in Fig. 3.

The simulation results and the protection method are unsuitable for use in microgrids. This is true even when mutual communication via the Goose protocol between the individual protectors and the switching of the different sets of settings is considered.

This is due to the low current ratio involved, and the time delay would also make designing the protection system much more complicated. In the case of microgrids, the design for topology change is very difficult.

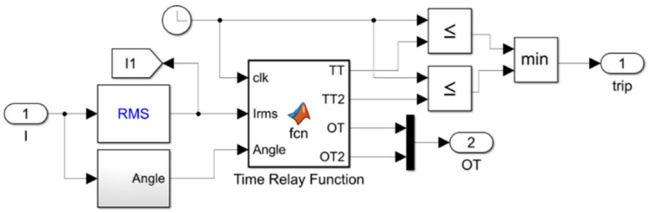


Fig. 3. Model diagram of overcurrent protection relay [7].

C. Directional overcurrent protection relay

Directional overcurrent protection is another option considered for line protection, which is also currently used for line protection. The schematic diagram of the directional overcurrent protection model for use for line protection is shown in Fig. 4.

The simulation results confirm the feasibility of using the system in practice for microgrids, although it is necessary to consider the mutual communication via Goose protocol between the individual protections for mutual blocking. The impact of the operation of this system will be influenced by the given topology and the location of the sources in each network.

The disadvantage of using directional overcurrent protection is the need for more powerful hardware and the availability of communication infrastructure.

The protection system used for the photovoltaic plant connected to Substation 3 offers both options, i.e., both directional overcurrent and differential, suitable for protecting the renewable resource.

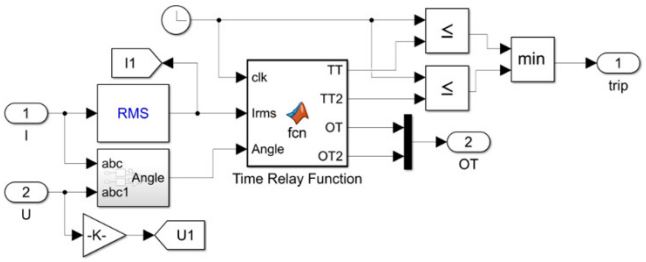


Fig. 4. Model diagram of directional overcurrent protection relay [7].

IV. Conclusion and Further Research

Designing a new protection system for use in microgrids is very challenging. It needs to be approached with sensitivity. If only renewable energy sources are considered, the value of the short-circuit current will be low due to the silicon (current) limit of the inverter. Although new power transistors can produce higher short-circuit currents, their value will still be around 1.3- to 2-times the nominal current. This makes the design even more complex, while various studies have shown that microgrid design using only renewable energy sources is not possible due to their dependence on weather effects. Therefore, the aim of this research is to demonstrate the need for energy mix designs of sources that can supply the area even in pure island mode in all weather conditions. In addition, this research provides proven protection functions that can be used to further develop the protection system. Future research will focus on fault location on the delineated line, taking only these two tested protection functions into account. To better advance the development of protection systems to be used in the future, it is necessary to further analyze the operating parameters for different scenarios and different energy mixes.

Acknowledgment

This research was funded by the Slovak Research and Development Agency [APVV-19-0576] and [APVV-21-0312] and the Ministry of Education, Science, Research and Sport of the Slovak Republic and the Slovak Academy of Sciences [VEGA 1/0757/21].

References

[1] R. Stefko et al., "Prospective Direction of Development of Protection Systems for Microgrids, " 2022 22nd International Scientific Conference on Electric Power Engineering (EPE), Kouty nad Desnou, Czech Republic, 2022, pp. 1-5.

[2] R. Stefko et al., "Case Study of Power Plants in the Slovak Republic and Construction of Microgrid and Smart Grid, " in *Appl. Sci.*, 11th ed. vol. 11, Basel (Switzerland), 2021, pp. 1-22.

[3] R. Stefko et al., "Výstavba a vývoj mikrosietí vo svete, " in *Elektroenergetika: International Scientific and Professional Journal on Electrical Engineering: Medzinárodný vedecký a odborný časopis pre elektroenergetiku*, 15th ed. vol. 1, Košice (Slovakia), 2022, pp. 16-19.

[4] R. Stefko et al., "Problems of protection of industrial networks with a high share of renewable energy sources, " 2020 IEEE 3rd International Conference and Workshop in Óbuda on Electrical and Power Engineering (CANDO-EPE), Budapest, Hungary, 2020, pp. 79-84.

[5] R. Stefko, "Protection system for microgrid and smart grid systems, " 21st Scientific Conference of Young Researchers (SCYR), Košice, Slovakia, 2021, pp. 23-26.

[6] R. Stefko et al., "Case Study of the Design of Renewable Energy Sources for Microgrid Systems, " 11th International Scientific Symposium ELEKTROENERGETIKA (EE), Stará Lesná, Slovakia, 2022, pp. 235-239.

[7] R. Stefko et al., " Modeling of Protection Relays and Renewable Energy Sources for Microgrid Systems, " in *Acta Electrotechnica et Informatica*, 22nd ed. vol. 3, Košice (Slovakia), 2022, pp. 9-17.

Compressed sensing of bio-signals

[1]*Jozef KROMKA (2nd year),*
Supervisor: [2]*Ján ŠALIGA*

[1,2]Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

[1]jozef.kromka@tuke.sk, [2]jan.saliga@tuke.sk

***Abstract*—This paper briefly introduces the use of Compressed Sensing (CS) techniques for the acquisition of bio-signals, specifically the Electrocardiogram (ECG) and pulse wave signals. The author presents two methods developed during his postgraduate research. The first method utilizes Multiwavelet functions and a unique approach to CS to acquire ECG signals. The second method employs CS to measure the impulse response of the wrist to obtain a pulse wave signal inside the radial artery, which can then be utilized for blood pressure (BP) estimation.**

***Keywords*—Compressed sensing, Bio-signals, Electrocardiogram, Pulse wave, Blood pressure**

I. Introduction

Cardiovascular diseases (CVD) are the most common cause of death worldwide [1]. One of the ways how to prevent CVD is through continuous long-term monitoring of cardiovascular parameters such as BP, heart rate, cardiac output, ECG, and others [2]. However, standard methods often rely on bulky and uncomfortable instruments that are not well-suited for continuous long-term monitoring of a patient. One solution to this problem is the use of wearable devices that are small, and lightweight, for continuous long-term monitoring. These wearable devices should be designed to minimize power consumption and maximize monitoring time and battery lifespan.

CS [3] is a technique that can be used to overcome these challenges. Traditional compression algorithms typically acquire signals according to the sampling theorem, followed by a transform to sparsify the signal and store or transmit only the most important coefficients. However, this approach can be power-intensive, making it unsuitable for use in wearable devices. In contrast, CS enables compression to be performed directly during signal acquisition, without the need for additional calculations.

II. Multiwavelet-based CS of ECG

In recent years, several CS methods for acquiring ECG signals have been proposed [4], [5]. However, these methods have limitations, such as the requirement for patient-specific information or the use of a trained dictionary. Additionally, some of these methods involve an additional step of detecting the QRS complex. Therefore, there is a need for a CS-based ECG acquisition method that does not require patient-specific information, a QRS complex detector, or a trained dictionary.

A new method, introduced by the author in [6], addresses the limitations of previously mentioned methods. The block diagram of this method is shown in Fig. 1. The ECG signal is acquired at sampling frequency f_s and split into frames with length N. Then a DC offset is removed from the frame. The resulting signal inside the frame is divided into low-frequency

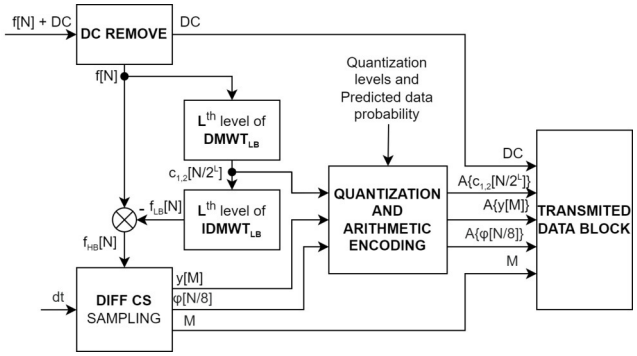


Fig. 1. Proposed method of ECG CS

and high-frequency components using Multiwavelet filters. Multiwavelet filtering was performed using a discrete Multiwavelet transform (DMWT) and inverse DMWT (IDMWT) from the Multiwavelet toolbox [7]. The low-frequency component is quantized and encoded using arithmetic encoding and stored in a data block. A novel way of CS acquires the high-frequency component. This novel way of CS is called Samples Difference Thresholding (SDT). SDT performs CS by comparing the two following samples and if the difference is higher than the defined constant dt the later sample is stored in the compressed measurement signal. The block which performs SDT also constructs the sensing matrix. The compressed measurement signal is quantized and subsequently encoded by arithmetic encoding. The measurement matrix is only encoded by arithmetic encoding.

The reconstruction of the ECG signal is performed in a simple way. Firstly all data are decoded and dequantized. Then the high-frequency signal component is reconstructed by Multiwavelet basis and CS reconstruction method. At the end of the reconstruction DC offset, high, and low-frequency components are added together forming a reconstructed signal.

The method was evaluated in terms of percentage root mean squared difference (PRD) and compression ratio (CR). The results for seven Multiwavelets are shown in Fig. 2. From the results, it can be seen that the BAT01 and DB2 Multiwavelet achieved the best PRD to CR ratio. With the following information, the next experiment focused on the

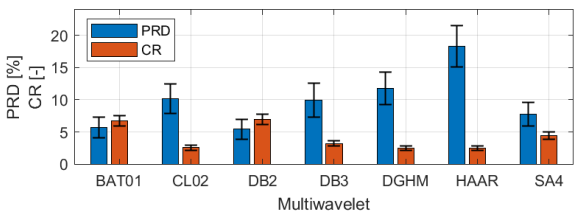


Fig. 2. PRD and CR results for 7-bit quantization

influence of the quantization on the PRD and CR after the reconstruction. From this experiment, it was concluded that the method could operate at 6-8 bits per sample instead of 11 bits and achieve good reconstruction quality with keeping a high CR. The details about the theory and performed simulation and case study can be found in [6].

III. Blood Pressure Estimation Using CS

Recent advancements in BP measurement have been made in [8], [9]. Some of these methods involve the use of wearable electrodes that measure the pulse wave in the radial artery to estimate BP. However, these methods typically require the use of different frequencies and a calibration step. Therefore, there is a need for a technique that can sense the pulse wave using a broadband signal, which would be more efficient and convenient for practical use.

The presented measurement method [10] is based on the determination of the impulse response of the radial artery on the wrist by exploiting a CS-based technique, followed by the estimation of the pulse wave propagation signal from the maximum values of the obtained impulse responses for each frame (i.e., time stamp). This work was developed during the author's stay at the University of Sannio in Benevento, as a result of an international collaboration with this institution.

The block diagram of the wrist impulse response estimation is outlined in Fig. 3.

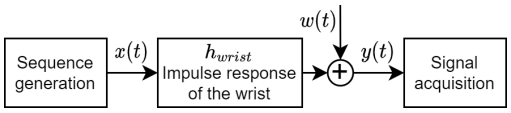


Fig. 3. The proposed measurement method

A pseudo-random binary sequence, $x(t)$, is employed as a broadband signal for the purpose of determining the impulse response of the wrist, $h_{wrist}(t)$. The acquired signal, $y(t)$, which is obtained by convolving $x(t)$ and $h_{wrist}(t)$, is also influenced by additive white Gaussian noise, $w(t)$.

To accurately capture the acquired signal, $y(t)$, the sampling frequency must be at least twice the maximum frequency present in the broadband signal, $x(t)$. However, since the impulse response, $h_{wrist}(t)$, is sparse in the time domain, the requirements for the analog-to-digital converter and memory can be reduced by using CS techniques.

In order to simulate impulse response $h_{wrist}(t)$ a 3D impedance model of the wrist was developed in MATLAB and simulated in LTSpice. The illustration of the developed 3D model is shown in Fig. 4.

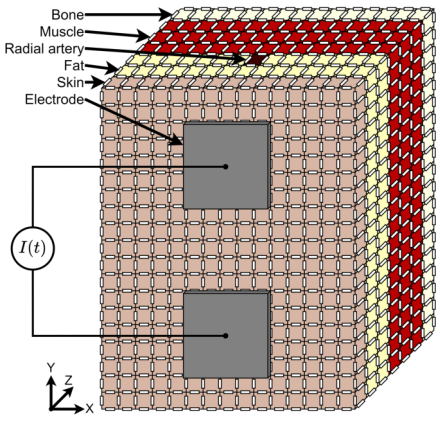
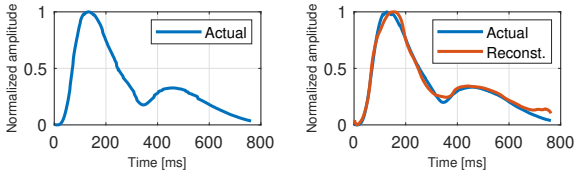


Fig. 4. The developed 3D model of wrist

The generated impulse response of the wrist was then used to perform simulation of the proposed method. The results for pulse wave estimation is shown in Fig. 5.



a) b)

Fig. 5. a) actual pulse wave propagation, b) reconstruction of pulse wave propagation with $CR = 2, L = 640, PRD = 7.954\%$.

The estimated and reconstructed pulse wave propagation signal shown in Fig. 5 can be used to calculate the pulse transmit time (PTT), which can be used as an estimator for BP. To compute PTT, data from two pulse wave propagation signals of blood circulation in the radial artery, along with their known distance, is required. The proposed measurement method can be extended by adding two more Bio-Z electrodes placed at a known distance to acquire this information. Detailed info about the method and performed simulation can be found in [10].

IV. Conclusion and Future Works

This paper presented a short overview of two novel methods for the acquisition of ECG and pulse wave signals developed during the author's postgraduate research. Further research on these methods is ongoing, with plans for future extension. The extended versions of these methods [6], [10], previously presented at the IMEKO TC-4 international conference, have been invited and already submitted for publication in the special issue of Measurement journal, a highly regarded journal in the field.

Acknowledgment

The work is a part of the project supported by the Science Grant Agency of the Slovak Republic (No. 1/0413/22).

References

[1] R. Jagannathan, S. A. Patel, M. K. Ali, and K. Narayan, "Global updates on cardiovascular disease mortality trends and attribution of traditional risk factors," *Current diabetes reports*, vol. 19, no. 7, pp. 1–12, 2019.

[2] G. Parati, G. S. Stergiou, E. Dolan, and G. Bilo, "Blood pressure variability: clinical relevance and application," *The Journal of Clinical Hypertension*, vol. 20, no. 7, pp. 1133–1137, 2018.

[3] D. L. Donoho, "Compressed sensing," *IEEE Transactions on information theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[4] L. De Vito, E. Picariello, F. Picariello, S. Rapuano, and I. Tudosa, "A dictionary optimization method for reconstruction of ECG signals after compressed sensing," *Sensors*, vol. 21, no. 16, 2021.

[5] J. Šaliga, I. András, P. Dolinský, L. Michaeli, O. Kováč, and **Kromka, Jozef**, "ECG compressed sensing method with high compression ratio and dynamic model reconstruction," *Measurement*, vol. 183, Oct. 2021.

[6] **Kromka, Jozef**, O. Kovac, J. Saliga, and L. Michaeli, "Multiwavelet-based ecg compressed sensing with samples difference thresholding," *25th IMEKO TC-4 International Symposium on Measurement of Electrical Quantities*, p. 215 – 220, 2022.

[7] **Kromka, Jozef**, O. Kovac, and J. Saliga, "Multiwavelet toolbox for matlab," *2022 32nd International Conference Radioelektronika*, 2022.

[8] T. Panula, J.-P. Sirkiä, D. Wong, and M. Kaisti, "Advances in non-invasive blood pressure measurement techniques," *IEEE Reviews in Biomedical Engineering*, vol. 16, pp. 424–438, 2023.

[9] Y. Yu, G. Anand, A. Lowe, H. Zhang, and A. Kalra, "Towards estimating arterial diameter using bioimpedance spectroscopy: A computational simulation and tissue phantom analysis," *Sensors*, vol. 22, no. 13, 2022.

[10] **Kromka, Jozef**, J. Saliga, O. Kovac, L. De Vito, F. Picariello, and I. Tudosa, "A novel cs-based measurement method for radial artery pulse wave estimation," *25th IMEKO TC-4 International Symposium on Measurement of Electrical Quantities*, p. 282 – 287, 2022.

Diagnostics of the surface of the contaminated insulator using the dielectric loss factor

[1]*Ľuboš Šárpataky (3rd year)*
Supervisor: [2]Bystrík Dolník

[1,2]Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic

[1]lubos.sarpataky@tuke.sk, [2]bystrik.dolnik@tuke.sk

***Abstract*—Insulators are a fundamental part of distribution and transmission lines. Although insulators are cheap components compared to other parts of transmission and distribution lines, their failure can lead to significant financial losses. Priority is to keep isolators in good condition to avoid the isolator's failure. Pollution of insulators combined with adverse weather conditions causes most of the problems that lead to the failure of insulators. Researchers introduced many diagnostic methods to prevent surface flashover on polluted insulators. Measurement of dielectric loss factor to evaluate surface conditions shows promising results and is useable at various voltage levels or frequencies. Two different diagnostic methods using dielectric loss factor show potential as a diagnostic method for polluted insulators.**

***Keywords*—dielectric loss factor, humidity, insulator, pollution**

I. Introduction

Insulators, a fundamental part of transmission and distribution lines, must meet some requirements. The essential parameters are electrical strength, mechanical strength, and surface conductivity. An increase in surface conductivity is usually the main reason for insulator failure. Insulators material and the environment influence the surface conductivity the most. All insulating materials (porcelain. glass, composites) have advantages and disadvantages. The selection of material depends mainly on prevailing environmental impacts. Pollution is the main factor that affects the surface conductivity of insulators. We know several types of pollution that are typical for different locations. We choose the correct type and material of insulators based on the pollution type [1]-[3].

Different methods use a different approaches to diagnose insulators. The fundamental diagnostics method is visual inspection, which requires qualified personnel. Other methods that do not require contact with the insulator are acoustic emission, infrared thermography, and night vision. In the last few years, power engineers started using drones to speed up the inspection process of large areas remotely [2]. Other methods use various measured quantities. The most often used is leakage current. Leakage current measurement can evaluate the condition of the insulator surface and has shown positive results with contamination identification. It increases with increasing pollution and with increasing humidity and thus can indicate the insulator's surface degradation. With sufficient data, the software can evaluate critical leakage current and a maintenance plan for the given location designed [4]-[7]. Research also deals with the harmonic analysis of leakage current to estimate contamination. Total harmonic distortion (THD) increases as insulator contamination increases [8]-[10]. They also evaluate individual harmonic components, mainly 3rd, 5th, and 7th [11], [12]. Numerical methods began to be used for diagnosis. Simulations in Finite Element Method (FEM), Boundary Element Method (BEM), and Finite Difference Method (FDM) can create a model whose results are very similar to experimental measurements [13]-[17].

The dielectric loss factor is already commonly used to diagnose transformers, generators, and cables [18], [19]. In practice, its spectroscopy is used, which points to different types of damage at different frequencies. For the diagnosis of insulators, the dielectric loss factor was used only in a few cases. It was mainly for the insulator's surface aging evaluation, where the results pointed to a change in the dielectric loss factor due to the aging of the insulating material [20]-[23]. However, its potential for diagnosing insulator contamination is promising based on the studied literature and our measurements.

II. Dielectric loss factor measurements

We use two different measurement methods of dielectric loss factor to evaluate the contamination of insulators. Firstly, we measure at low voltage in a wide frequency range with an LCR meter. The next step was the dielectric loss factor measurement using high voltage levels. For both measurements, we use porcelain insulators.

A. Low voltage measurement of the dielectric loss factor

To measure the dielectric loss factor at a 5 V voltage level and frequency range from 1 Hz to 200 kHz, we shorten the creepage distance of the porcelain insulator U 160 BL by using measuring electrodes made from copper tape to 4 cm. An impedance analyzer HIOKI LCR meter IM3533-1 directly evaluates the dielectric loss factor. We prepare four solutions representing four pollution levels (L1, L2, L3, L4). We make individual solutions and evaluate equivalent salt deposit density (ESDD) and non-soluble deposit density (NSDD) according to IEC/TS 60815-1:2008 standard to determine the contamination of the insulator.

Fig. 1 shows a significant increase in the dielectric loss factor when pollution levels increase. Rapid dielectric loss factor increase is visible between pollution levels L1 and L2. At low frequencies, up to 30 Hz, differences between measurements are not stepwise, and lower pollution level has higher dielectric

loss factor value in some measured points. At higher frequencies, after curves reach their peaks, the dielectric loss factor increases according to the pollution level. Curve peaks or rapid decreases are moving to higher frequencies depending on the pollution level. From all measurements done on various insulator types, we conclude that the measurement method is applicable to evaluate the contamination of insulators. The most significant differences are in higher frequencies which can help to avoid interference with the industrial frequency [7].

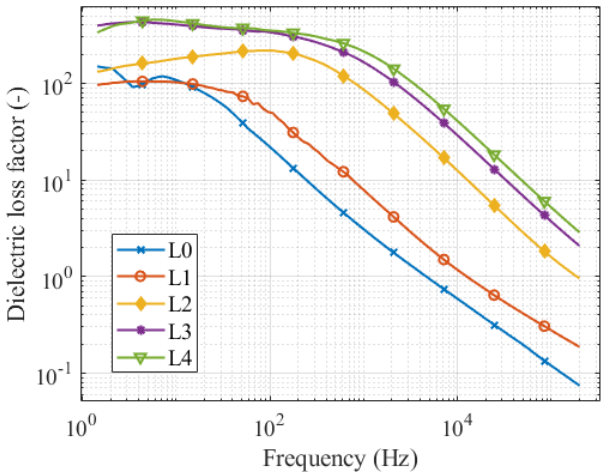


Fig. 1 Dielectric loss factor measured on a clean and polluted insulator at a relative humidity of 80%.

B. High voltage measurement of the dielectric loss factor

An Omicron MI 600 measuring system designed to measure capacitance, leakage current, and dielectric loss factor was used for measurements at high voltage levels. We measured and evaluated the dielectric loss factor and leakage current on the porcelain insulator at 5 kV, 7.5 kV, and 10 kV voltage. We contaminated the insulator with two different levels of pollution. These pollution levels had the same contamination parameters as the second and fourth pollution levels from previous measurements.

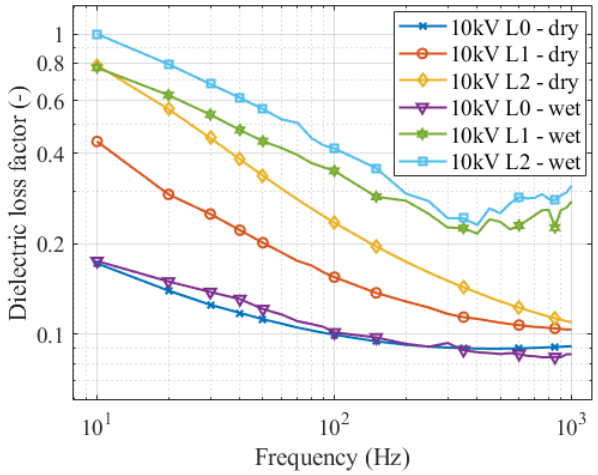


Fig. 2 Dielectric loss factor measured on a clean and polluted insulator with dry and wet pollution layer.

Measurement results show that the dielectric loss factor is more sensitive to changes in the conditions of the insulator surface than the leakage current. Changes in the level of pollution showed a more significant increase for both dry and wet conditions.

As Fig. 2 shows, the dielectric loss factor in dry conditions was more stable (with a lower standard deviation), with clearly identifiable progress. The advantage is that dry conditions are more suitable for diagnostics than wet ambient conditions. Compared with measurements at low voltage, the dielectric loss factor at high voltage was pronounced in lower frequencies. The dielectric loss factor is less sensitive to voltage changes than leakage current [24].

III. Conclusion and Further Research

The measurement of the dielectric loss factor at different voltages (low and high) showed that it is usable for evaluating and diagnosing insulator contamination. The dielectric loss factor is used for various devices in power engineering to monitor aging. Pollution of insulators has a similar effect to aging. Both increase surface conductivity. Our measurement results show that pollution increases the dielectric loss factor. Compared with leakage current, the dielectric loss factor grows faster, and the changes are more significant and noticeable. An important finding is dielectric loss factor stability when the voltage changes [24]. In further research, we will focus on the industrial frequency of 50 Hz and evaluate the applicability of the dielectric loss factor in a wide range of voltages (from 5 kV to 22 kV) and on several types of insulators. We will use disk (cap and pin) and long rod insulators. We will also compare glass, porcelain, silicone rubber, and ethylene-propylene-diene monomer (EPDM) insulating materials used for insulators. All measurements will be done on clean insulators as well as on four levels of pollution. Another variable will be relative humidity. It will vary from room conditions (around 40%) to 90% relative humidity.

Acknowledgment

This research was funded by the Scientific Grant Agency of the Ministry of Education, Science, Research, and Sport and of the Slovak Academy of Sciences under contract No. 1/0340/18; the Slovak Research and Development Agency under contract No. APVV-15-0438; and the Cultural and Educational Grant Agency of the Ministry of Education, Science, Research, and Sport of the Slovak Republic (KEGA) under project No. 008TUKE-4/2019.

References

[1] Farzaneh, M.; Chisholm, W.A. Insulators for Icing and Polluted Environments; IEEE Press: Piscataway, NJ, USA; John Wiley & Sons: Hoboken, NJ, USA, 2009.

[2] Vosloo, W.L.; Macey, R.E.; de Tourreil, C. The Practical Guide to Outdoor High Voltage Insulators; Crown Publications: Johannesburg, South Africa, 2004.

[3] Haddad, A.; Warne, D.F.; Institution of Electrical Engineers. (Eds.) Advances in High Voltage Engineering; Institution of Electrical Engineers: London, UK, 2004.

[4] Chatterjee, S.; Banik, A.; Dalai, S.; Chatterjee, B. Identification of salt and salinity level of 11kV contaminated porcelian disc insulator using STD-MRA analysis of leakage current. In Proceedings of the 2015 International Conference on Condition Assessment Techniques in Electrical Systems (CATCON), Bangalore, India, 10–12 December 2015; pp. 242–246. https://doi.org/10.1109/CATCON.2015.7449543.

[5] Dey, J.; Dutta, S.; Baral, A.; Chakravorti, S. Leakage Current Monitoring of Suspension Insulator for Effective Determination of ESDD. In Proceedings of the 2019 8th International Conference on Power Systems (ICPS), Jaipur, India, 20–22 December 2019; pp. 1–6. https://doi.org/10.1109/ICPS48983.2019.9067562.

[6] Haya, S.N.; Jaelani, A.A.; Lumba, L.S.; Suwarno. Accelerated Aging Study on Silicone Rubber Insulator with Contaminant Level and Fog Conductivity Variations. In Proceedings of the 2021 3rd International

Conference on High Voltage Engineering and Power Systems (ICHVEPS), Bandung, Indonesia, 5–6 October 2021; pp. 91–95. https://doi.org/10.1109/ICHVEPS53178.2021.9601100.

[7] Dolník, B.; Šárpataky, Ľ.; Kolcunová, I.; Havran, P. Sensing Method Using Multiple Quantities for Diagnostic of Insulators in Different Ambient Conditions. Sensors 2022, 22, 1376. https://doi.org/10.3390/s22041376.

[8] Putra, N.R.M.; Sartika, N.; Rachmawati; Suwarno. The study on leakage current waveform characteristics and computer simulation of ceramic insulator under artificial tropical condition. In Proceedings of the 2018 12th International Conference on the Properties and Applications of Dielectric Materials (ICPADM), Xi'an, China, 20–24 May 2018; pp. 320–323. https://doi.org/10.1109/ICPADM.2018.8401273.

[9] Deb, S.; Ghosh, R.; Dutta, S.; Dalai, S.; Chatterjee, B. Condition monitoring of 11kV porcelain pin insulator extracting surface current from total leakage current. In Proceedings of the 2017 3rd International Conference on Condition Assessment Techniques in Electrical Systems (CATCON), Rupnagar, India, 16–18 November 2017; pp. 403–406. https://doi.org/10.1109/CATCON.2017.8280253.

[10] Wakhidin, M.; Suwarno. Effects of Artificial Tropical Climate Aging on Insulation Performance of Silicone Rubber Polymeric Insulators. In Proceedings of the 2019 2nd International Conference on High Voltage Engineering and Power Systems (ICHVEPS), Denpasar, Indonesia, 1–4 October 2019; pp. 1–6. https://doi.org/10.1109/ICHVEPS47643.2019.9011130.

[11] A. A. Salem, R. Abd-Rahman, S. A. Al-Gailani, M. S. Kamarudin, H. Ahmad, and Z. Salam, 'The Leakage Current Components as a Diagnostic Tool to Estimate Contamination Level on High Voltage Insulators', *IEEE Access*, pp. 1–1, 2020, DOI: 10.1109/ACCESS.2020.2993630.

[12] A. A. Salem, R. Abd Rahman, N. A. Mohd Jamail, N. Azlin Othman, I. Ullah, and H. B. Ahmad, 'New Approach for Monitoring Contamination Level on Outdoor Insulator Based on Harmonics Components of the Leakage Current', in *2021 IEEE International Conference on the Properties and Applications of Dielectric Materials (ICPADM)*, Johor Bahru, Malaysia, Jul. 2021, pp. 418–421. DOI: 10.1109/ICPADM49635.2021.9493975.

[13] Araya, J.; Montana, J.; Schurch, R. Electric Field Distribution and Leakage Currents in Glass Insulator Under Different Altitudes and Pollutions Conditions using FEM Simulations. *IEEE Latin Am. Trans.* **2021**, *19*, 1278–1285. https://doi.org/10.1109/TLA.2021.9475858.

[14] Sardast, R.; Faghihi, F.; Vakilian, M. Study on Different Dimensions of C-type Corona Rings in 400 kV Insulator Strings Based on FEM Analysis of Electric Field Distribution. In Proceedings of the 2019 27th Iranian Conference on Electrical Engineering (ICEE), Yazd, Iran, 30 April–2 May 2019; pp. 673–678. https://doi.org/10.1109/IranianCEE.2019.8786424.

[15] Zhou, X.; Hu, D.; Yin, J.; Wang, S. Study on Mechanical Properties of Fallen Suspension Porcelain Insulator using FEM Analysis. In Proceedings of the 2021 IEEE 5th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Xi'an, China, 15–17 October 2021; pp. 813–817. https://doi.org/10.1109/ITNEC52019.2021.9586970.

[16] Othman, M.; Isa, M.; Mazlee, M.N.; Piah, M.A.M.; Abd, N. Rahman. Simulation of 33kV String Insulators Using Finite Element Method (FEM). In Proceedings of the 2019 IEEE Student Conference on Research and Development (SCOReD), Bandar Seri Iskandar, Malaysia, 15–17 October 2019; pp. 86–89. https://doi.org/10.1109/SCORED.2019.8896273.

[17] Bhavani, J.; Kumar Ch, S.V.S.P. Finite Element Modeling of Voltage and Electric Field Distribution along the Insulators. In Proceedings of the 2019 4th International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT), Bangalore, India, 17–18 May 2019; pp. 225–229. https://doi.org/10.1109/RTEICT46194.2019.9016905.

[18] X. Cheng, G. Ye, H. Sun, T. Li, and C. Sun, "Analysis of low-frequency dielectric loss of XLPE cable insulation based on extended Debye model," *AIP Advances*, vol. 11, no. 8, p. 085103, Aug. 2021, DOI: 10.1063/5.0060939.

[19] S. Morsalin and B. T. Phung, "Modeling of dielectric dissipation factor measurement for XLPE cable based on Davidson-Cole model," *IEEE Trans. Dielect. Electr. Insul.*, vol. 26, no. 3, pp. 1018–1026, Jun. 2019, doi: 10.1109/TDEI.2019.007784.

[20] V. Yalentic, S. Grzinic, and D. Dobrec, "Testing the electrical insulation system of power transformer based on mesuring factor of dielectric losses," in *IEEE EUROCON 2017 -17th International Conference on Smart Technologies*, Ohrid, Macedonia, Jul. 2017, pp. 423–427. doi: 10.1109/EUROCON.2017.8011146.

[21] M. Jaroszewski, A. Beroual, and D. Golebiowski, "Effect of temperature on dielectric loss factor of biodegradable transformer oil," in *2018 IEEE International Conference on High Voltage Engineering and Application (ICHVE)*, ATHENS, Greece, Sep. 2018, pp. 1–4. doi: 10.1109/ICHVE.2018.8641763.

[22] N. Bashir, H. Ahmad, and M. Shaffuan Suddin, "Ageing studies on transmission line glass insulators using dielectric dissipation factor test," in *2010 Conference Proceedings IPEC*, Singapore, Singapore, Oct. 2010, pp. 1062–1066. doi: 10.1109/IPECON.2010.5696973.

[23] F. Gerdinand, M. Budde, a M. Kurrat, "Electrical and mechanical strength of mineral filled epoxy insulators in correlation to power loss factor", in *Proceedings of the 2004 IEEE International Conference on Solid Dielectrics, 2004. ICSD 2004.*, Toulouse, France, 2004, p. 320–323. DOI: 10.1109/ICSD.2004.1350355.

[24] Ľ. Šárpataky, B. Dolník, J. Zbojovský, U. Schichler, O. Pischler, and B. Schober, "Sensing Method Using Dielectric Loss Factor to Evaluate Surface Conditions on Polluted Porcelain Insulator," Sensors, vol. 22, no. 23, p. 9442, Dec. 2022, doi: 10.3390/s22239442.

Ontology for visualization tasks in data analytical processes

[1]*Juliana IVANČÁKOVÁ (4th year)*
Supervisor: [2]Peter BEDNÁR

[1,2]Department of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1]juliana.ivancakova@tuke.sk, [2]peter.bednar@tuke.sk

***Abstract*— This paper focuses on describing the ontology for data analytical processes and details one of the five modules that focuses on describing data visualization within these processes. The modules that have been created are extensions to existing ontologies that cover the phases of the CRISP-DM methodology. Such ontologies include, for example, OntoDM, DSO, DMOP, DMWF and the VISO ontology. The VISO ontology was the inspiration for the creation of a concept-specific module that describes not only the visualization of the data itself in the context of data understanding or preprocessing, but also the graphical visualization of models for data analytical processes.**

***Keywords*—Data Analysis, Ontology, Visualization, Semantic model**

I. Introduction

Visualization is an important aspect of data analysis as it allows for the representation of complex data sets in a way that is easy for humans to understand. Overall, visualization plays a crucial role in the data analytical process as it allows for the discovery of insights and patterns that may not have been apparent otherwise. The following section briefly describes the ontologies that cover data analytical processes. Section III focuses on the definition of the developed model for data analytical processes, which consists of 4 modules. Section IV is dedicated to a more detailed description of the concepts that have been defined for data visualization.

II. Coverage of data analytical processes by ontologies

There are several ontologies that partly follow data-analytical processes, but none of them are complete.

OntoDM [1] is a specific type of ontology that is used in the field of data mining. It is a formal representation of knowledge that is specific to a particular domain and is used to define and represent the concepts and relationships within a data set. The goal of OntoDM is to provide a common vocabulary and understanding of the data for data mining tasks such as classification, clustering, and association rule mining.

DSO [2] is a way of organizing and classifying the concepts and entities within the field of data science. It helps to define the relationships between different aspects of data science work and provides a framework for understanding and communicating about the field.

The DMOP Ontology [3] describes the internal mechanism of data mining, including the numerical optimization method, the error function and the learning rules. This ontology covers 3 phases of CRISP-DM. The DMWF ontology [4] defines data operators. These operators are used in the process of data pre-processing, modelling, and evaluation. The operators are described in a similar way to the services as their inputs, outputs, assumptions, and effects.

Visualization is also an important part of the data mining process. There is an ontology that also describes some data visualization concepts. VISO [5] is an ontology for describing visual scenes and objects in a structured way. It is designed to be a common language for different computer vision systems and applications to describe visual information. The ontology is organized into a hierarchy of classes and properties that describe various aspects of visual scenes, including objects, attributes, and relationships. Viso is distinguished by its 7 modules: Graphics - formalizing concepts related to graphical relationships and representations; Data - defines the variables and structures of the data; Facts - formalizes constraints, linkages and default values; Activity - deals with the human aspect within the visualization; System - this module covers both the HW and the SW; User - characterizes the user extensions; and Domain - describes the specifications of the domain.

All these ontologies cover just some of the data mining tasks, not all of them. Therefore, we created our own ontology that extends the existing ones with the necessary concepts and combines them into one complete ontology.

III. Model for data analytical processes

Based on existing ontologies, an ontology consisting of the following 4 modules was created [6] and gradually expanded.

A. Domain concepts

Domain formalization concepts are specified using the SKOS metamodel [7], which allows naming, describing and defining concepts localized in multiple natural languages. Concepts can be organized hierarchically in the form of a thesaurus/taxonomy using the sessions skos#broader/skos#narrower. Polyhierarchical concepts can also be defined as schemas. They can be associatively linked by the skos#related session in addition to the hierarchical arrangement. The defined common dictionary of domain terms can also be used as a classification scheme, a scheme for organizing different types of documents. These documents can either be used as domain documentation for data analysis, or they can be created by the data analyst to document the data and the results of the analysis process itself. The

document types themselves can be specified in SKOS as classifying taxonomies.

B. Data elements and Performance indicators

These are fundamental notions that depict input and output data and measurable performance metrics. To represent the data, the proposed semantic model introduces two basic classes: the logical data attribute representing the data - the DataElement class, and its physical implementation - the PhysicalDataElement class. Logical data elements are used for the description of any input data attribute that is identified by the domain expert as relevant for the solution of the given data analysis task, or for the description of any output data generated during the analysis process, including predictions, data analysis models, or values of measurable domains and technical performance indicators.

C. Algorithms and data analytical models

The Algorithm class, derived from the Operator class, is the main type for representing algorithms. The Operator class represents a general operation that takes an input of a type and converts it to the required output (i.e., in the case of an algorithm, a training set of data for a data analysis model). The class Model represents the output of the algorithm, i.e., the general data-analytical model. Some types of models can be directly applied to new data (for example, classification or regression models), so specific subclasses of the type Algorithm can be derived from the class Operator.

D. Process model for data analysis workflow

The proposed process model can be used to generate automated workflows for the task at hand, as well as to describe existing scripts and data analysis to make them replicable and reusable. The state is represented by the set of instances assigned to the shared variables, as in the semantic services model. It is composed of nodes that represent individual operations (Operator class) during data pre-processing, modelling and evaluation, or control modules (Control node class).

IV. VISUALIZATION IN DATA ANALYTICAL PROCESSES

Main contribution of this article is an extension of the data analysis model with concepts that describe visualization in data mining processes. The key concepts are Algorithm, which is a type of operator that depends on input variables, and VisualizationMethod, which is an operator that takes data or a model as input and outputs a visualization in the form of a graph.

The Fig. 1 illustrates the concepts of data visualization. As an example, we have selected the visualization of PDP method [8] for explaining machine learning models. At the beginning, it is important to define the type of visualization that is required. In principle, any visualization requires data as input. Either the data can be visualized as part of the data understanding or preprocessing phase, or the output of the visualization can be in the form of graphical representations of the models used for data mining. The concept of Algorithm is treated as a class operator, which is represented as a generic operation that transforms an input of a specific type into the desired output. The output of algorithms is then represented by the Model class. This concept is used together with data as input for the PDP (Partial Dependence Plot), a graphical tool that helps to understand the relationship between the input and the predicted variable. The outputs of the PDP operator are used to compute PDPStats, defined as TechnicalIndicator, which is a technical indicator specified based on domain indicators. The values obtained from PDPStats are inputs to the PDPVisualMethod. The final desired output is a PDP graph, which is a subclass of the VisualizationMethod operator and is specified in this case as a Graph.

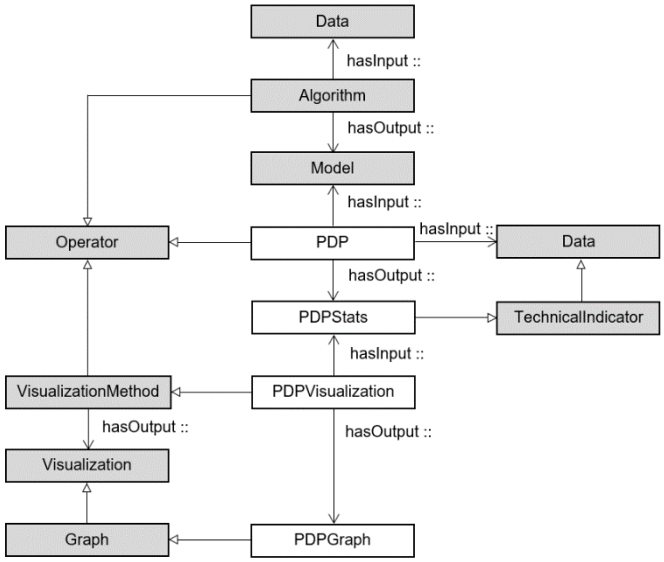


Fig. 1 Visualization in data analytical processes

V. CONCLUSION

Evaluation of the model will be carried out on several case studies focusing on data analysis and the evaluation will be based on selected code coverage metrics [9].

ACKNOWLEDGMENT

The research presented in this paper was financially supported by the grant APVV-16-0213 and VEGA 1/0685/21.

REFERENCES

[1] P. Panov, S. Dzeroski and L. N. Soldatova, "OntoDM: An Ontology of Data Mining", In: 2008 IEEE International Conference on Data Mining Workshops, 2008, pp. 752–760.

[2] E. Patterson, I. Baldini, A. Mojsilovic and K. Varshney, "What is the Data Science Ontology?", [online] [cit. 31.01.2023] doi: <https://www.datascienceontology.org/help.>.

[3] M. Hilario, P. Nguyen, H. Do, A. Woznica and A. Kalousis, "Ontology-Based Meta-Mining of Knowledge Discovery Workflows", In: Meta-Learning in Computational Intelligence, 2011.

[4] J. Vanschoren and L. Soldatova, "Exposé: An ontology for data mining experiments", In: International workshop on third generation data mining: Towards service-oriented knowledge discovery (SoKD-2010), 2010, pp. 31–46.

[5] J. Polowinski and M. Voigt, "VISO: a shared, formal knowledge base as a foundation for semi-automatic infovis systems", v CHI '13 Extended Abstracts on Human Factors in Computing Systems, Paris France, 2013, pp. 1791–1796, doi: 10.1145/2468356.2468677.

[6] P.Bednár, J.Ivančáková and M.Sarnovský. "Semantic automatiozation of the data-analytical processes", In: International Symposium on Applied Computational Intelligence and Informatics, 2022.

[7] A. Miles and S. Bechhofer, SKOS Simple Knowledge Organization System Reference, W3C Recommendation. World Wide Web Consortium, 2009, doi: https://www.w3.org/TR/skos-reference/.

[8] Ch.Molnar, "Interpretable Machine Learning", A Guide for Making Black Box Models Explainable, 2023, [online] [cit. 2023-03-02] doi: <https://christophm.github.io/interpretable-ml-book/>.

[9] S.Pittet, "What is code coverage?", [online] [cit. 2023-01-29] doi: <https://www.atlassian.com/continuous-delivery/software-testing/code-coverage>.

Dielectric properties evaluation of ester-based nanofluids with fullerene and magnetite nanoparticles

[1]*Miloš ŠÁRPATAKY (3rd year)*
Supervisor: [2]Juraj KURIMSKÝ

[1,2] Department of Electric Power Engineering, FEI TU of Košice, Slovak Republic

[1]milos.sarpataky@tuke.sk, [2]juraj.kurimsky@tuke.sk

***Abstract*—Nanofluid research is moving forward and therefore it is necessary to unify the results to find the optimal combination of the base fluid and nanoparticles of optimal concentration for the practical tests in high-voltage equipment. This paper determines the optimal concentration of fullerene and magnetite nanoparticles in natural and synthetic esters as base fluids. Evaluation of examined samples and optimal concentration is according to AC breakdown voltage, dissipation factor, relative permittivity, and volume resistivity measurements.**

***Keywords*—nanofluids, optimal concentration, natural ester, synthetic ester, breakdown voltage.**

I. Introduction

Nanofluids as a possible alternative to today's most frequently used insulating liquids (mainly mineral oils) for high voltage equipment are undergoing intensive research [1]. The goal of this research in general is to find suitable, stable, economically and environmentally acceptable alternatives that meet the requirements of insulating and heat transfer medium [2]. Research of nanofluids focuses on thermal properties (for heat transfer) or dielectric properties (quality of insulation) that will be described in this paper. In general nanoparticles in base, oils enhance their heat transfer and dielectric properties and this improvement has its peak at some level of concentration of nanoparticles in base fluid [3]. The optimal concentration of nanoparticles is mostly determined by one experiment for example by AC breakdown voltage test [4], [5], lightning breakdown voltage test [6], [7], partial discharge investigation [8], [9] etc., therefore, determination of optimal concentration is less objective. This paper focuses on optimal concentration determination and evaluation of samples made of natural and synthetic esters with fullerene (3 concentrations) and magnetite (2 concentrations) nanoparticles. For evaluation of samples AC breakdown voltage test and dielectric performance (dissipation factor, relative permittivity and volume resistivity) at temperatures from 25°C to 140°C.

II. Experimental methods

AC breakdown voltage test was performed in accordance with IEC60156 by Haefely Hipotronics AC test set. Mushroom-like electrodes 2.50±0.05 mm apart with a 12.5 mm diameter were horizontally positioned in a test cell. The volume of the test cell was 350 ml. Voltage rise with a rate of 2 kV/s and measurement of each sample was executed 30 times. A more detailed description of this experiment is in [10].

The measurement of thermal dependence dielectric performance was executed according to IEC 60247 2004. As a test cell, HAEFELY TEST AG 2903 H was used with a volume of 40 ml. Tettex 2830 a Precision Oil and Solid Dielectric Analyzer was used for the analysis of dielectric performance quantities at different temperatures from 25°C to 140°C with a step of 10°C [11].

For the evaluation of individual experiments performance ranking is created. Nanofluids or pure liquids with the highest quality among all samples got a low-rank number starting with number 1 (from Table II to V). For the determination of optimal concentration and evaluation of liquid insulation quality, a sum of rankings was made in Table VI. Lower ranking values mean higher sample quality. The mean value for thermal dependence tests (tanδ, relative permittivity and volume resistivity) was calculated as an average value of all examined temperatures. The results of these experiments were published in [10], [11]. Concentrations and sample names are in Table I.

TABLE I
EXAMINED SAMPLES

Base oil	Nanoparticles	Concentration	Sample name
NE	-	-	NE
NE	C_{60}	0.01%w/V	1_NE
NE	C_{60}	0.02%w/V	2_NE
NE	C_{60}	0.03%w/V	3_NE
NE	Fe_3O_4	0.05%w/V	1_NE_FE
NE	Fe_3O_4	0.1%w/V	2_NE_FE
SE	-	-	SE
SE	C_{60}	0.01%w/V	1_SE
SE	C_{60}	0.02%w/V	2_SE
SE	C_{60}	0.03%w/V	3_SE
SE	Fe_3O_4	0.05%w/V	1_SE_FE
SE	Fe_3O_4	0.1%w/V	2_SE_FE

III. Experimental results

A. AC breakdown voltage

In Table II there are the results of the AC breakdown voltage test with mean value (average value of 30 measurements) and change (the percentage difference between values of the

nanofluid and its base fluid). Positive change means that nanofluid reached a higher average value of AC breakdown voltage and negative value showed the opposite result. A more detailed description of the results is in [10].

TABLE II
RESULTS OF AC BREAKDOWN VOLTAGE TEST

Sample	Mean BDV (kV)	Change (%)	Performance ranking (-)
2_SE_FE	65.06	13.52	1
1_NE_FE	63.81	14.68	2
1_SE_FE	63.79	11.31	3
3_NE	58.57	5.27	4
SE	57.31	-	5
2_NE_FE	57.18	2.77	6
3_SE	56.71	-1.06	7
2_NE	56.36	1.29	8
NE	55.64	-	9
2_SE	55.06	-4.09	10
1_SE	53.22	-7.69	11
1_NE	34.61	-60.76	12

B. Dissipation factor (tanδ)

Since the dissipation factor is related to the power loss, the lowest possible number is required. The negative change, in this case, represents better quality of insulating material (Table 3). Samples 1_NE_FE and 2_NE_FE with higher temperatures changed their properties and the dielectric analyser displayed an error message connected with overcurrent, so the measurement can not be finished. The values at low examined temperatures for both samples were around 4.6.

TABLE III
RESULTS OF THE DISSIPATION FACTOR THERMAL DEPENDENCE TEST

Sample	tanδ (-)	Change (%)	Performance ranking (-)
3_NE	0.00721285	-16.94	1
NE	0.00843454	-	2
2_NE	0.00894131	6	3
1_NE	0.00994308	17.88	4
2_SE	0.04412792	-96.57	5
3_SE	0.04850869	-78.82	6
1_SE	0.06613369	-31.16	7
SE	0.08674085	-	8
2_SE_FE	0.37111923	328	9
1_SE_FE	0.37249769	329	10
1_NE_FE	N/A	-	11
2_NE_FE	N/A	-	11

C. Relative permittivity

Relative permittivity test results are in Table 4. Same to the dissipation factor, the relative permittivity in insulation materials requires a low value. Samples 1_NE_FE and 2_NE_FE can not be measured at all temperatures for the same reason of overcurrent. However, from low-temperature results, it is obvious that sample 1_NE_FE with a value of around 3.2 got better results than sample 2_NE_FE with a value of over 5.

TABLE IV
RESULTS OF RELATIVE PERMITTIVITY THERMAL DEPENDENCE TEST

Sample	Relative permittivity (-)	Change (%)	Performance ranking (-)
2_NE	2.88930769	-1.21	1
1_NE	2.89123077	-1.14	2
3_NE	2.90807692	-0.56	3
NE	2.92423077	-	4
1_SE_FE	3.05423077	-0.04	5
SE	3.05561538	-	6
2_SE	3.05607692	0.01	7
1_SE	3.05646154	0.02	8
3_SE	3.06938462	0.45	9
2_SE_FE	3.07038462	0.48	10
1_NE_FE	N/A		11
2_NE_FE	N/A		12

D. Volume resistivity

The results of volume resistivity are in Table 5. Insulation quality rise with rising resistivity. Sample 2_NE_FE can not be measured due to a system error of the equipment.

TABLE V
RESULTS OF VOLUME RESISTIVITY THERMAL DEPENDENCE TEST

Sample	Volume resistivity (GΩm)	Change (%)	Performance ranking (-)
NE	206.186923	-	1
3_NE	202.545385	-1.8	2
1_NE	136.708462	-50.82	3
2_NE	135.803846	-51.83	4
2_SE	36.9943077	173.9	5
3_SE	33.8745385	150.6	6
1_SE	21.9004615	62.2	7
SE	13.5023554	-	8
1_NE_FE	9.52769231	2064	9
2_SE_FE	2.44823077	-451.5	10
1_SE_FE	2.39284615	-464.3	11
2_NE_FE	N/A		12

IV. DISCUSSION AND CONCLUSIONS

The overall performance ranking shows the samples with the highest quality among examined samples. The quality decrease with a higher value of performance ranking sum. The overall results are in Table VI.

TABLE V
RESULTS OF VOLUME RESISTIVITY THERMAL DEPENDENCE TEST

Sample	Overall performance ranking (-)	Overall ranking avg (-)	Performance ranking overall value (-)
3_NE	10	6.00	1
2_NE	16	10.67	2
NE	16	11.33	3
1_NE	21	15.00	4
SE	27	12.33	5
2_SE	27	15.67	6
3_SE	28	14.00	7
1_SE_FE	29	11.67	8
2_SE_FE	30	10.67	9
1_NE_FE	33	12.33	10
1_SE	33	18.33	11
2_NE_FE	41	17.67	12

For a more objective classification of the AC breakdown voltage measurement the overall statistics column "Overall ranking avg" represent the sum of breakdown voltage test rank and an average number of thermal dependence tests. The rank of samples according to the overall result is in "Performance ranking overall value".

Results showed that the optimal concentration of fullerene nanoparticles in NE is 0.03%w/V and in SE is 0.02%w/V. NE and its nanofluids with fullerene nanoparticles reached the best results among all samples. According to this research, SE is negatively influenced by both kinds of examined nanoparticles and magnetite nanoparticles in both esters decreased its quality. For a more objective evaluation of samples, the lightning breakdown voltage and partial discharge measurement will be part of the overall research of this set of nanofluids.

Acknowledgment

This research was funded by the Ministry of Education, Youth and Sports within the project VEGA 2/0011/20 and 1/0154/21 and the Slovak Agency for Research and Development based on contracts no. APVV-15-0438, APVV-17-0372, and APVV-18-0160.

References

[1] RAFIQ, M. et al. Use of vegetable oils as transformer oils – a review. In *Renewable and Sustainable Energy Reviews* . 2015. Vol. 52, s. 308–324. .

[2] RAFIQ, M. et al. Transformer oil-based nanofluid: The application of nanomaterials on thermal, electrical and physicochemical properties of liquid insulation-A review. In *Ain Shams Engineering Journal* . 2020. s. S2090447920301805. .

[3] ŠÁRPATAKY, M. et al. Dielectric Fluids for Power Transformers with Special Emphasis on Biodegradable Nanofluids. In *Nanomaterials* . 2021. Vol. 11, no. 11, s. 2885. .

[4] BEROUAL, A. - KHALED, U. Effect of Nanoparticles' Mixtures on AC Breakdown Voltage of Mineral Oil. In *IEEE Transactions on Dielectrics and Electrical Insulation* . 2021. Vol. 28, no. 4, s. 1216–1222. .

[5] KHELIFA, H. et al. Effect of Conducting, Semi-Conducting and Insulating Nanoparticles on AC Breakdown Voltage and Partial Discharge Activity of Synthetic Ester: A Statistical Analysis. In *Nanomaterials* . 2022. Vol. 12, no. 12, s. 2105. .

[6] BEROUAL, A. - KHALED, U. Statistical Investigation of Lightning Impulse Breakdown Voltage of Natural and Synthetic Ester Oils-Based Fe3O4, Al2O3 and SiO2 Nanofluids. In *IEEE Access* . 2020. Vol. 8, s. 112615–112623. .

[7] KHALED, U. - BEROUAL, A. Lightning impulse breakdown voltage of synthetic and natural ester liquids-based Fe3O4, Al2O3 and SiO2 nanofluids. In *Alexandria Engineering Journal* . 2020. Vol. 59, no. 5, s. 3709–3713. .

[8] KHELIFA, H. et al. AC Breakdown Voltage and Partial Discharge Activity in Synthetic Ester-Based Fullerene and Graphene Nanofluids. In *IEEE Access* . 2022. Vol. 10, s. 5620–5634. .

[9] JACOB, J. et al. Partial Discharge Characteristics of Nanofilled Mineral Oil. In *2019 IEEE Region 10 Symposium (TENSYMP)* . 2019. s. 316–320. .

[10] ŠÁRPATAKY, M. et al. Synthetic and natural ester-based nanofluids with fullerene and magnetite nanoparticles – An experimental AC breakdown voltage study. In *Journal of Molecular Liquids* . 2022. Vol. 368, s. 120802. .

[11] ŠÁRPATAKY, M. et al. Dielectric Performance of Natural- and Synthetic-Ester-Based Nanofluids with Fullerene Nanoparticles. In *Energies* . 2023. Vol. 16, no. 1, s. 343. .

Crossover of encryption algorithms

[1]*Peter Pekarčík (1st year),*
Supervisor: [2]*Eva CHOVANCOVÁ*

[1,2]Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

[1]peter.pekarcik@tuke.sk, [2]eva.chovancova@tuke.sk

***Abstract*—In today world, people use a lot of data in the digital form. They store it to devices, manipulate with them or send it to other devices. People often use computers and smartphones to store and transfer confidential data. This data should be available only to authorized people. A proven way how to reach this is using of encryption algorithms. There exist dozens of algorithms. Some of them were broken and today are not considered as safe. We assume, that with the trend of increasing computer performance, sooner or letter there will be a need to develop new encryption algorithms. One of the method of a creation of new algorithms is using crossover technique. This paper describes the actual state of using crossover for a creation of encryption algorithms. Moreover, there are described two known examples of crossover - modified Playfair cipher and modified substitution cipher approach. Our goal is to summarize the current knowledge in the field of crossing encryption algorithms, so that we can use this technique to develop new, better algorithms.**

***Keywords*—crossover, encryption, encryption algorithm, fitness function, mutation, selection**

I. Introduction

Data security is one of the most prospective field of the modern data science. The reason why it is so is, that people are using more and more data in the digital form. If data contain confidential information, we have to ensure inaccessibility to unauthorized people. A proven method, how to reach that is using of encryption algorithms. Today, encryption is used in various fields to ensure confidentiality. UAV communication is one of these fields. Most of the most used encryption algorithms were invented between years 1976 and 1998. For example, DES was invented in 1976, RSA in 1977, 3DES in 1981, Blowfish in 1993 and Twofish in 1998.[1][2] We emphasize, that since 1998 significant progress has been made in the field of IT. Moreover, we expect this progression will be continuing. Therefore, we are asked questions: Are these algorithms sufficient to protect data even today? DES is still a strong code and has never been broken, although several high-powered computers are now available which, using brute force, can crack the code.[3] The most used encryption algorithm, RSA, should be broken by a quantum computer around a year 2035.[4] Don't new, more robust algorithms have to be developed? Do we have to start developing new algorithms all over again, or is it enough to improve already developed algorithms? The most popular encryption techniques are based on mathematical operations like shifts, permutations or substitutions. We expect, that we can use some parts of older algorithms, and by combining them to create new, better algorithms. Developing of the new algorithms will be beneficial only if provided solutions, that will be more secure, faster, more effective or less memory intensive. One of the well-know methods for development of new algorithms is a crossover. But can we use the crossover to create new algorithms that meet any of these assumptions?

II. Actual knowledge

As was mentioned in the introduction, our goal is to create new algorithms, that will be better in some way. If we want to achieve this result, we have to proceed systematically. We chose the crossover as a means of the developing.

Studies tell us, that Genetic algorithms for cryptography are better than typical algorithms. Genetic algorithms contain processes/operations, which are bioinspired, such as:

1) **selection of Fitness function** - Determination of the formula will be used to evaluate the quality of the result.
2) **selection** - Section of algorithms suitable for crossover.
3) **crossover** - Creating a new algorithm based on the choice in the selection process.
4) **mutation** - Editing of the created algorithm in step 3, so that the result is in the expected form.

They are used to optimizing and generating high-quality solutions.[5][6] The sequence of these steps used for the encryption is shown in the following diagram:

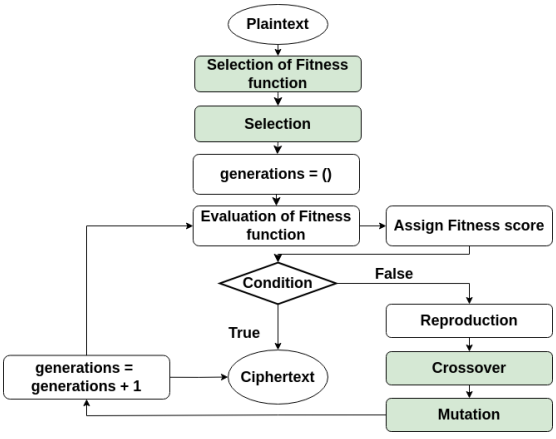


Fig. 1. Encryption process[7]

The first process is **selection of the Fitness function**. The role of the Fitness function is representing the requirements the population should adapt to meet. It forms the basis for selection, and so it facilitates improvements. It represents the task to be solved in the evolutionary context. In the example, if the task is to find an integer x that maximizes x^2, the fitness of genotype 10010 could be defined by decoding its corresponding phenotype ($10010 \rightarrow 18$) and then taking its square: $18^2 = 324$.[8]

The goal of fitness function is to evaluate the quality of the result. In our case of encryption algorithms, should be constructed to achieve improvement of:

1) **security** - In order to achieve better security of the created algorithm will be assembled series of tests programs. They role will be verifying resistance to different types of attacks. For example, in case RSA the most known attacks are[9]:
   a) Low decryption exponent attack against RSA
   b) Partial key exposure attack
   c) Chosen ciphertext attack against RSA
   d) Common modulus attack against RSA

   It is necessary to pay attention to the fact that attacks are tied to specific type of encryption algorithm. After the crossover has been performed, it will be necessary to examine how the crossover has affected the resistance to that type of attack.
   As the safest encryption standard is considered AES.[10]
2) **speed** - Every algorithm will be tested for encryption speed. The reason is, that speed of encryption is one of the most important factor for encryption. Testing will be done using standard Linux commands like[11]:
   a) `time`
   b) `clock()`
   c) `date`

   As the fastest encryption algorithm is considered Blowfish followed by DES and Triple DES.[12]
3) **memory requirements** - During development of new encryption technique, we have to check memory requirements of algorithm. If the amount of the used memory will be high, the algorithm will be unusable. Our goal is to create algorithms with less or equal memory requirements or as algorithms entering the process of a crossover. On Linux, we can use[11]:
   a) instrumentation framework Valgrind
   b) command `time`, e.g.:`$ time -v exeFile`
   c) command `/usr/bin/time`, e.g.:
      `$ /usr/bin/time -v exeFile`

   As the encryption algorithm consumes, the least memory usage is considered AES.[13]

Our goal will be to develop algorithms, that will be preferable in the worst case in one of the mentioned and in the best case in all of them.

The second process is **selection of algorithms for the crossover**. Today exists dozens of different encryption algorithms. Examples of the most used are[14]:

1) Data Encryption Standard (DES)
2) Triple Data Encryption Standard (3DES)
3) Advance Encryption Standard (AES)
4) Blowfish
5) RSA

But are all of these algorithms suitable for crossover? Every of these algorithms is different. However, we can look at every of them as a sequence of steps. In the literature, that has been studied, we have not found any restrictions for selection algorithms inputted to crossover. Can we do crossover on the arbitrary algorithm? In this case, we can not forget about division of encryption algorithms into symmetric and asymmetric algorithms. In symmetric key encryption, any user using the encryption system has a copy of the single secret key, this secret key is used for encryption and decryption, it is faster than asymmetric encryption. In symmetric encryption, the sender share the same secret key with the receiver to use in the decryption process. Since the secret key is shared with the sender and the receiver, it becomes risky to use it. Asymmetric key encryption involves multiple keys for encryption. [15][16]

Before the beginning of the solving the question, if is any of algorithms appropriate for a crossover, we have to find out what features of algorithms to focus on. Of course, we can test every possible combination of algorithms by the trial and error method, but this approach is not very scientific. If we try every combination of existing encryption algorithms, it would take a lot of time and our result would not be comprehensive. This is the reason why, we have to examine a suitability of every encryption algorithm for crossover. In the currently available sources regarding the crossing of algorithms, there are no mentions of the unsuitability of some encryption algorithms for crossover. One of the relevant method how to reach it, is detail analysis of each algorithm. We plan to analyze algorithms for suitability on crossover.

After that, we can continue to the **crossover**. When we require done this step precisely, we have to divide each encryption algorithm to the separate parts. Here comes another question: How to divide algorithm to the separate steps automatically? We assume, that dividing into independent steps will be solvable using an appropriate machine learning algorithm. But what algorithms are suitable for this purpose?

Mentioned steps help us to work on result of crossover. This will be useful only, if the obtained algorithm will be correct and more effective as algorithms on which the crossing process was performed. As we mention above, we are going to analyze more factors for evaluation of the obtained result. Doesn't it happen that improving of one factor (e.g. encryption speed) causes another factor to deteriorate? Which of these factors are more important for us - security, speed or memory requirements? To what extent are we able to accept the deterioration of one of the factors at the expanse of another? However, we must not forget that the result will be usable only if the ciphers obtained with it can also be decrypted. So, in addition to a new encryption algorithm, a new decryption algorithm is also needed.

The last process of crossover is **mutation**. This is to prevent falling all population into a local optimum of solved problem. Mutation changes randomly the offspring.[7] We expect, that in the crossing process we will work with a string in a form suitable to a computer but not for a human (e.g. binary form or hexadecimal form). We have to transfer this form to the more suitable form (e.g. ascii form).

Now that we have enough information about the crossover, we can look at the currently known algorithms created by this process. The first example is a novel method using crossover with matrix technique to modifying cipher by using Playfair. Our goal is encrypting text *DAVINCHI* by modifying Playfair cipher. Steps of the modified Playfair cipher are[17]:

1) duplicated letters are replaced by the letter *X*:
   DAVINCHX
2) letters are divided to the pairs:
   DA VI NC HX
3) we chose the first pair of letters:
   DA
4) we denoted for **DA** like this:

$D = 1\ 0\ 0\ 1\ 0\ 0\ 1\ 1\ 1 \quad A = 1\ 0\ 1\ 1\ 1\ 0\ 1\ 1\ 1$

5) we choose a point for crossover:
$D = 1\ 0 \mid 0\ 1 \mid 0\ 0\ 1\ 1\ 1 \quad A = 1\ 0 \mid 1\ 1 \mid 1\ 0\ 1\ 1\ 1$
6) we make crossover (swap between bits) we get results:
$D \rightarrow F = 1\ 0\ 0 \mid 1\ 1 \mid 0\ 1\ 1\ 1$
$A \rightarrow M = 1\ 0\ 1 \mid 1\ 0 \mid 0\ 1\ 1\ 1$
7) we chose the second pair of letters:
VI
8) we denoted for **VI** like this:
$V = 1\ 0\ 1\ 1\ 0\ 1\ 1\ 0\ 0 \quad I \rightarrow J = 1\ 0\ 0\ 1\ 0\ 1\ 1\ 1\ 0$
9) we choose a point for crossover:
$V = 1\ 0 \mid 1\ 1 \mid 0\ 1\ 1\ 0\ 0 \quad J = 1\ 0 \mid 0\ 1 \mid 0\ 1\ 1\ 1\ 0$
10) we make crossover (swap between bits) we get results:
$V \rightarrow U = 1\ 0 \mid 0\ 1 \mid 0\ 1\ 1\ 0\ 0$
$I \rightarrow R = 1\ 0 \mid 1\ 1 \mid 0\ 1\ 1\ 1\ 0$
11) we chose the third pair of letters:
NC
12) we denoted for **NC** like this:
$N = 0\ 1\ 1\ 1\ 0\ 0\ 1\ 0\ 0 \quad C = 0\ 1\ 1\ 1\ 0\ 1\ 1\ 0\ 1$
13) we choose a point for crossover:
$N = 0\ 1\ 1\ 1\ 0 \mid 0\ 1 \mid 0\ 0 \quad C = 0\ 1\ 1\ 1\ 0 \mid 1\ 1 \mid 0\ 1$
14) we make crossover (swap between bits) we get results:
$N \rightarrow O = 0\ 1\ 1\ 1\ 0 \mid 1\ 1 \mid 0\ 0$
$C \rightarrow Z = 0\ 1\ 1\ 1\ 0 \mid 0\ 1 \mid 0\ 1$
15) we chose the fourth pair of letters:
HX
16) we denoted for **HX** like this:
$H = 1\ 1\ 1\ 1\ 0\ 0\ 1\ 0\ 0 \quad X = 1\ 1\ 1\ 1\ 0\ 1\ 1\ 0\ 1$
17) we choose a point for crossover:
$H = 1\ 1\ 1\ 1\ 0 \mid 0\ 1 \mid 0\ 0 \quad X = 1\ 1\ 1\ 1\ 0 \mid 1\ 1 \mid 0\ 1$
18) we make crossover (swap between bits) we get results:
$H \rightarrow G = 1\ 1\ 1\ 1\ 0 \mid 1\ 1 \mid 0\ 0$
$X \rightarrow E = 1\ 1\ 1\ 1\ 0 \mid 0\ 1 \mid 0\ 1$
19) in new generation, for H we have G, and for X we have E. The plaintext *DAVINCI* will be encrypted as *FMUROZGE*
20) we choose a key word for PlayFair matrix to cipher it again as shown:

N	O	V	E	L
M	T	H	D	A
B	C	F	G	I
K	P	Q	R	S
U	W	X	Y	Z

21) then we get the cipher text as the rules of PlayFair method and its condition will be:
$(FM\ UR\ OZ\ GE) = (HB\ KY\ WL\ RD)$

The second example is a modified substitution cipher approach. Our goal is encrypting text *World'14*[18]:

1) text is placed to the spiral from point $[0, 0]$ clockwise
2) any non-prime random number is taken as the first key:
$$FirstKey = 3160420$$
3) is computed:
$$R = Firstkey \ mod\ 256 \quad (1)$$
$$R = 3160420\ mod\ 256 = 100$$
4) the substitution for the plaintext is given: In the 1^{st} round of spiral, substitution for: 'W', 'o', 'r', 'l':

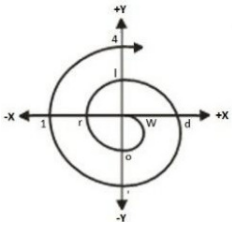


Fig. 2. Spiral co-ordinate for Encryption[18]

$ASCII(W) = 87 \Rightarrow 87 + 100 = 187 = ASCII(╗)$
$ASCII(o) = 111 \Rightarrow 111 - 101 = 10 = ASCII(LF)$
$ASCII(r) = 114 \Rightarrow 114 - 102 = 12 = ACSII(\ ♀)$
$ASCII(l) = 108 \Rightarrow 108 + 103 = 211 = ASCII\ (╙)$
In the 2^{nd} round, we continue on the first internal spiral path, substitution for 'd', "', '1', '4':
$ASCII(d) = 100 \Rightarrow 100 + 104 = 204 + ASCII(W) = (204 + 87)\ mod\ 256 = 35 = ASCII(\#)$
$ASCII(') = 111 \Rightarrow 39 - 105 = -66 + ASCII(o) = (-66 + 111)\ mod\ 256 = 45 = ASCII(-)$
$ASCII(1) = 49 \Rightarrow 49 - 106 = -57 + ASCII(r) = (-57 + 114)\ mod\ 256 = 57 = ACSII(9)$
$ASCII(4) = 52 \Rightarrow 52 + 107 = 159 + ASCII(4) = (159 + 108)\ mod\ 256 = 11 = ASCII(♂)$
5) now we get string ╗ LF ♀ ╙ # − 9 ♂
6) we transform the string to the ASCII values and this value on binary 8-bits string:
$╗ => ASCII(187) = ASCII(10111011)$
$LF => ASCII(10) = ASCII(00001010)$
$♀ => ASCII(12) = ASCII(00001100)$
$╙ => ASCII(221) = ASCII(11011101)$
$\# => ASCII(35) = ASCII(00100011)$
$- => ASCII(45) = ASCII(00101101)$
$9 => ASCII(57) = ASCII(00111001)$
$♂ => ASCII(11) = ASCII(00001011)$
7) we divide a binary ASCII string of each character and perform *Hybrid crossover*. Let chose as pivot second number in the binary string and let's sign operation of crossover as X. Then:
For $ASCII(10111011)$ and $ASCII(00001010)$:
1011 X 1010 1011 X 0000
1010 1011 1000 0011
10101000 XX 00111011
10111000 00101011
For $ASCII(00001100)$ and $ASCII(11010011)$:
0000 X 0011 1100 X 1101
0011 0000 1101 1100
00110000 XX 11011100
00000001 11111100
For $ASCII(00100011)$ and $ASCII(00101101)$:
0010 X 1101 0011 X 0010
0001 1110 0010 0011
00010010 XX 00111110
00111110 00010010
For $ASCII(00111001)$ and $ASCII(00001011)$:
0011 X 1011 1001 X 0000
0011 1011 1000 0001
10101000 XX 00111011
10111000 00101011
8) now, we have final result:

$$ASCII(10111000) = ASCII(184) = ╕$$
$$ASCII(00101011) = ASCII(43) = +$$
$$ASCII(00000001) = ASCII(1) = ☺$$
$$ASCII(11111100) = ASCII(252) = \eta$$
$$ASCII(00111110) = ASCII(62) =>$$
$$ASCII(00010010) = ASCII(18) =↕$$
$$ASCII(00011000) = ASCII(24) =↑$$
$$ASCII(00101011) = ASCII(59) =;$$

Therefore, the final cipher text is: $╕ + ☺\eta >↕↑;$

III. SUMMARY OF SOLVED AND UNSOLVED PROBLEMS

Two mentioned examples present actual known algorithms created by crossover. In the first mention algorithm, was modified PlayFair encryption and in the second was modified substitution cipher approach. There exist couple examples of algorithms created by crossover. Probably, when we would study next papers, we will find next examples of new encryption techniques. For example, in the article Multilevel Approach for Cryptography using Genetic Algorithms with Existing S-DES Key Generation Method is presented a multilevel approach for cryptography using the S-DES Key Generation method and genetic algorithm. Result is, that convergence of S-DES and Genetic Algorithm is more secure and unbreakable than traditional cryptosystem. In the article, authors mentioned, that they would like to implement and upgrade the performance of genetic algorithms and experiments for different advance ciphers like AES or DES.[19]

In the next article Modification on RSA cryptosystem using genetic optimization, authors introduce a modified ciphering technique. The created system is a combination of symmetrical system using Genetic algorithm Inspired Cryptography and asymmetrical system using RSA cryptography. The feature of such approach includes high data security and high feasibility for practical implementation.[20]

We know from both examples that there are attempts at development of new encryption algorithms using the crossover. Thus, we assume that it makes sense to consider the suitability of other encryption algorithms for crossover.

IV. CONCLUSION

In this paper, we present a novel approach to creation of new encryption algorithms. We summarize the most important current knowledge in the field of cryptography. At the end of summarizing, we concluded the need for new encryption algorithms. We chose crossover as the method of developing these algorithms. Next, we described the actual knowledge in the field of crossover of algorithms. Part of this summarizing was also description of two known algorithms created by crossover method - modifying Playfair cipher and modified substitution cipher. We mention also next algorithm S-DES, but we did not attend to its description. We are planning to investigate the suitability of next mentioned algorithms for crossover and creation of new encryption algorithms by this method.

ACKNOWLEDGMENT

This publication was realized with support of the Operational Programme Integrated Infrastructure in frame of the project: Intelligent systems for UAV real-time operation and data processing, code ITMS2014+: 313011V422 and co-financed by the Europen Regional Development Found.

REFERENCES

[1] R. Rivest, A. Shamir, and L. Adleman, "A methond for obtaining digital signitures and public-key cryptosystems," 1978. [Online]. Available: https://people.csail.mit.edu/rivest/Rsapaper.pdf

[2] Fergustron and et al., "The skein hash function family," 2010. [Online]. Available: https://www.schneier.com/wp-content/uploads/2015/01/skein.pdf

[3] W. J. Buchanan, 2010. [Online]. Available: https://www.researchgate.net/publication/228952572_3DES_Encryption_and_Decryption_in_Microsoft_NET

[4] I. F. Muller and M. van Heesch, "Migration to quantum-safe cryptography," 2020. [Online]. Available: https://publications.tno.nl\T1\guilsinglrightTNO-2020-migration

[5] P. Srikanth, A. Mehta, N. Yadav, S. Singh, and S. Singhal, "Encryption and decryption using genetic allorithm operations and pseusorandom numbers," vol. 118, 2017. [Online]. Available: http://ijcsn.org/IJCSN-2017/6-3/Encryption-and-Decryption-Using-Genetic-Algorithm-Operations-and/-Pseudorandom -Number.pdf

[6] "A study on cryptography using gernetic algorithm," *International Journal of Computer Applications*, vol. 118, May 2015. [Online]. Available: https://research.ijcaonline.org/volume118/number20/pxc3903559.pdf

[7] S. Lavkush, K. P. Bhupendra, and S. Ramgopal, "Breaking of Simplifiet Data Encryption Standard Using Genetic Algorithm," *GLOBAL JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY*, vol. 12, p. 57, March 2012. [Online]. Available: https://www.semanticscholar.org/paper/Breaking-of-Simplified-Data-Encryption-Standard-Sharma-Pathak/5fc04935ba6658ea5baa28d58be7b7ebf9ac9335

[8] A. E. Eiben and J. Smith, *Introduction to Evolutionary Computing, Second Edition*, ser. Natural Computing Series. Springer, 2015. [Online]. Available: https://warin.ca/ressources/books/2015_Book_IntroductionToEvolutionaryComp.pdf

[9] A. H. Lone, "Commnon attacks on rsa and its variants with possible countermeasures," 2016. [Online]. Available: https://www.researchgate.net/publication/316588561_Common_Attacks_on_RSA_and_its_Variants_with_Possible_Countermeasures

[10] J. S. Kiran, "Cryptography: The sciene of secure communication," vol. 16, 2016. [Online]. Available: https://www.researchgate.net/publication/348147932_Cryptography_The_Sciene_of_Secure_Communication

[11] W. Shott, *The Linux Command Line*, ser. A LinuxCommand.org Book. No Strach Press, 2013. [Online]. Available: https://wiki.lib.sun.ac.za/images/c/ca/TLCL-13.07.pdf

[12] M. Panda, "Performance evaluation of symmetic encryption algorithms for information security," 2017. [Online]. Available: https://www.researchgate.net/publication/261429690_Performance_analysis_of_encryption_algorithms_for_Information_Security

[13] O. Alshannaq, "Analysis of the lowes memory consumption (memory usage) through running different cryptiography techniques for different types of images," 2022. [Online]. Available: https://iopscience.iop.org/article/10.1088/1742-6596/2319/1/012027/pdf

[14] "Performence analysis of data encryption algorithms," 2006. [Online]. Available: https://www.cse.wustl.edu/~jain/cse567-06/ftp/encryption_perf.pdf

[15] V. K. Sourbah Katoch, Sumit Singh Chauhan, "A raview on genetic algorithm: past, present, and future," 2020. [Online]. Available: https://link.springer.com/article/10.1007/s11042-020-10139-6

[16] A. Humadi, "Encryption," 2020. [Online]. Available: https://iopscience.iop.org/article/10.1088/1742-6596/2319/1/012027/meta

[17] M. Mohammed Sami, "Novel method using crossover (genetic algorithms) with matrix tqchnique to modifying ciphering by using playfair," vol. 3, 2013. [Online]. Available: https://djes.info/index.php/djes/article/view/512/472

[18] S. K. Mahata, S. Nigaja, S. Srivastava, M. Dey, and S. Som, "A novel approach to cryptography using modified substitution cipher and hybrid crossover technique," *International Jouenal of Computer Applications*, 2014. [Online]. Available: https://www.researchgate.net/publication/306262581_A_Novel_Approach_to_Cryptography_using_Modified_Substitution_Cipher_and_Hybrid_Crossover_Technique/citations

[19] D. P. Bagane, D. D. Dharrao, and D. S. Kotrappa, "Multilevel approach for cryptography using genetic algorithms with existing s-des key generation method," *International Journal of INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERINGs*, 2022. [Online]. Available: https://ijisae.org/index.php/IJISAE/article/view/2343

[20] "Modification on rsa crytosystem using genetic optimization," 2014. [Online]. Available: https://www.arpapress.com/Volumes/Vol19Issue2/IJRRAS_19_2_03.pdf

A Simple Stepper-Based Controller

[1]*Pavol SMOLEŇ (5th year)*
Supervisor: [2]František ĎUROVSKÝ

[1]BWG, k.s. (member of Redex Group), Prešov, Slovak Republic
[2]Dept. of Electrical Engineering and Mechatronics, FEI TU of Košice, Slovak Republic

[1]pali.smolen@gmail.com, [2]frantisek.durovsky@tuke.sk

Abstract **– The paper presents a stepper-based controller operating on the principle of step-increasing or step-decreasing of the reference. It is suitable for complex tasks requiring the control of system interaction with the environment by either controlling the interaction force or tension, forcing the system to comply with the motion of the environment.**

The paper describes the principle of the controller and practical implementation - controlling the force between two rolls of coating machine widely used in steel and aluminium industry.

Keywords **- stepper-based controller, coating machine, position control**

I. INTRODUCTION

Controllers of auxiliary drives play an important role in the industry. Typical examples are drives for folding of rolls for reduction of the bending in the strip [1] or rollers for application of the chemical treatment on the strip surface known as coating machines [2]. The controller design mainly relies on good system knowledge and identification or estimation of mechanical parameters and, regularly, parameters of the controller are obtained by trial-and-error methods [3], [4]. Complex systems usually require the involvement of fuzzy controllers [5].

A custom-made solutions, like the one proposed in our paper can be used in various applications. It is advantageous to use the described controller in cases when operational intervention of the controller allows a slow change of the controlled variable or when the time of the system response is unknown or when this time depends on various other parameters (roll surface, used technology) [6]. The position controller described in [7] can be used in combination with proposed stepper-based controller as shown in section practical implementation.

II. PRINCIPLE OF CONTROLLER

The stepper-based type of controller has been initially introduced in [8] for the control of tension in the annealing oven. Here, stepper-based controller is modified and used for the force control of the coating machine. Proposed stepper-based force controller is shown in Fig. 1.

It works on the principle of step-increasing or step-decreasing of the reference, in that case position, where following parameters are used:

- IN_1 - the 1st interval of force value in %,
- IN_2 - the 2nd interval of force value in %,
- t_1 - the pulse time for the 1st interval in *ms*,
- t_2 - the pulse time for the 2nd interval in *ms*,
- x_{add} - the step size in *mm*.

A flowchart of the proposed stepper-based controller can be found in Fig. 2. The stepper-based controller represents an infinite loop with the active position controller. If the difference ΔF, expressed as a percentage of the force setpoint, exceeds the interval set by IN_2, a direction of the next move is chosen and a new position setpoint x_{Fset} is calculated. The next step decides how long will stepper controller remain inactive after the new position setpoint is sent to the position controller. This is the time period intended for the mechanic system for appropriate response. The selection of parameters IN_1, IN_2, t_1 and t_2 must fulfill the conditions: $IN_1 > IN_2$ and $t_1 < t_2$.

III. PRACTICAL IMPLEMENTATION

The stepper-based force controller was experimentally verified in the coating machine what is one of the main component of the strip coating lines. The coating machine is designed to apply a thin layer of a liquid medium via rotating rolls on one or both surfaces on a running strip. The coater equipment is made of these main elements:

- a pan, which is filled with the coating medium,
- a pick-up-roll (PR) which scoops the liquid out of the pan and then transports it to the applicator-roll,
- an applicator-roll (AR) which applies the coating medium on the metal surface.

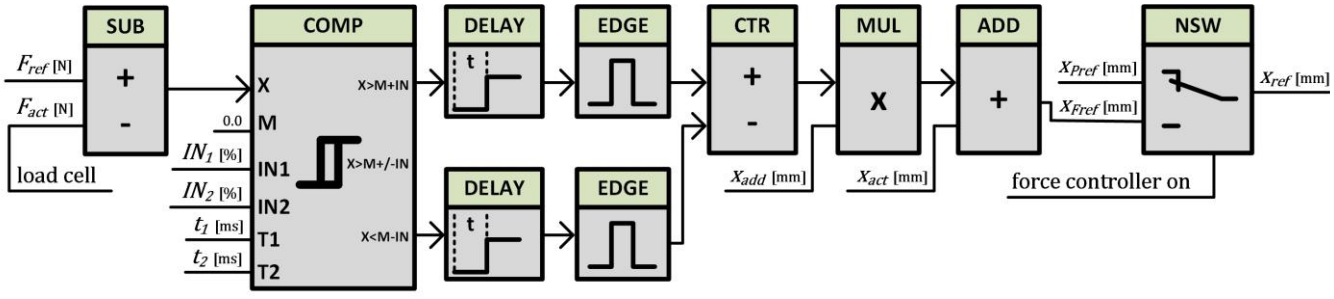


Fig. 1. Detailed block scheme of the stepper-based controller

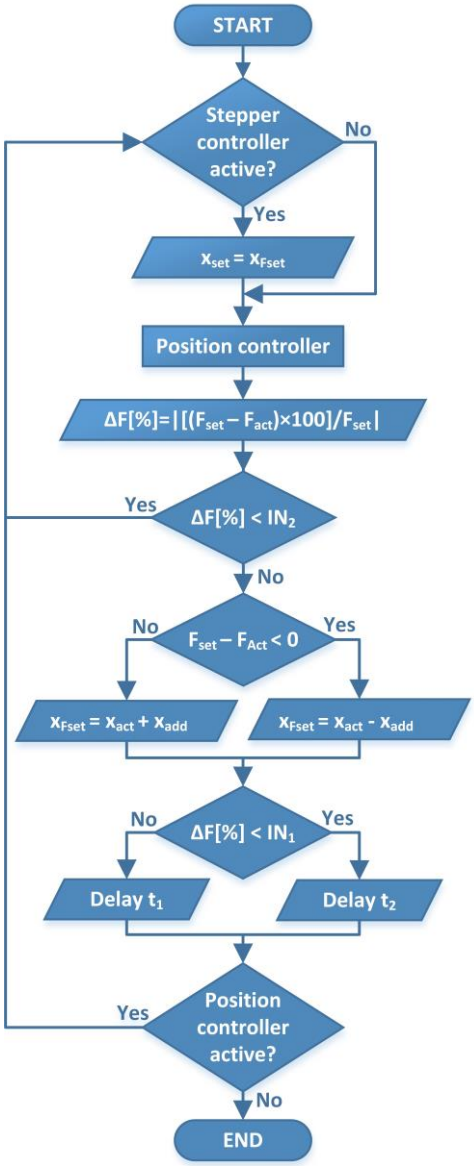


Fig. 2. The algorithm implemented in the stepper-based force controller

The position of the coater rolls is adjustable. Servo drives equipped with gearboxes, self-locking drive spindles and pressure feedback via load cells are utilized. The AR and PR rotate with a given speed, direction, gap and force between both rolls during coating process. Experimental results for combination of stepper-based force and position controller are shown in Fig. 3.

When the PR achieved a small gap with AR given as $x_{ref} = 0{,}5mm$, the stepper-based force controller is enabled while position controller remains active (time 14:18:03). The stepper controller is running to reach the force setpoint $F_{ref} = 500N$ with following parameters: $IN_1 = 30\%$, $t_1 = 1000_{ms}$, $IN_2 = 10\%$, $t_2 = 1500ms$, $x_{add} = 0{,}05mm$. It takes a few seconds until PR tips the AR and the actual force F_{act} starts increasing. When the percentage difference between F_{ref} and F_{act} goes below value given in IN_2, the force setpoint is reached. After that, both position and force controller remain active.

IV. Conclusion

The stepper-based controller has been presented in this paper. Accuracy can be adjusted by parameters given in percentage of the setpoint force. Controller was successfully applied on several applications in industry.

References

[1] Kornilov, G.P.; Abdulveleev, I.R.; Khramshin, T.R.; Shokhin, V.V. Advanced electric drive control system of continuous hot-dip galvanizing line. In Proceedings of the Proceedings - 2020 International Conference on Industrial Engineering, Applications and Manufacturing, ICIEAM 2020,

[2] Yamada, T.; Fujisaki, K. Application of linear induction motor for tension supply and heating to thin sheet steel. Electrical Engineering in Japan 2009, 168, 38–47.

[3] Hackl, C.M. Non-identifier Based Adaptive Control in Mechatronics; Springer: Cham, 2017.

[4] Jankowska, K.; Dybkowski, M. A Current Sensor Fault Tolerant Control Strategy for PMSM Drive Systems Based on Cri Markers. Energies 2021,

[5] He, F., Wang, Q.: Compensation and fuzzy control of tension in strip winding control system. In: IEEE International Conference Industrial Electronics and Applications, IEEE-ICIEA'12, Singapore (2012)

[6] Sabanovic, A.; Ohnishi, K. Motion Control Systems; John Wiley & Sons: Singapore, 2017

[7] Pavol Smoleň, A Position Controller with Low-Speed Area based on Switching of Non-linear Functions, 22nd Scientific Conference of Young Researchers, Kosice (Slovakia), pp. 79-80, ISBN 978-80-553-4061-6

[8] Magura, D.; Fedák, V.; Sanjeevikumar, P.; Kyslan, K. Tension Controllers for a Strip Tension Levelling Line. In Proceedings of the Advances in Systems, Control and Automation. Lecture Notes in Electrical Engineering; Springer: Singapore, 2018; Vol. 42, pp. 33–44.

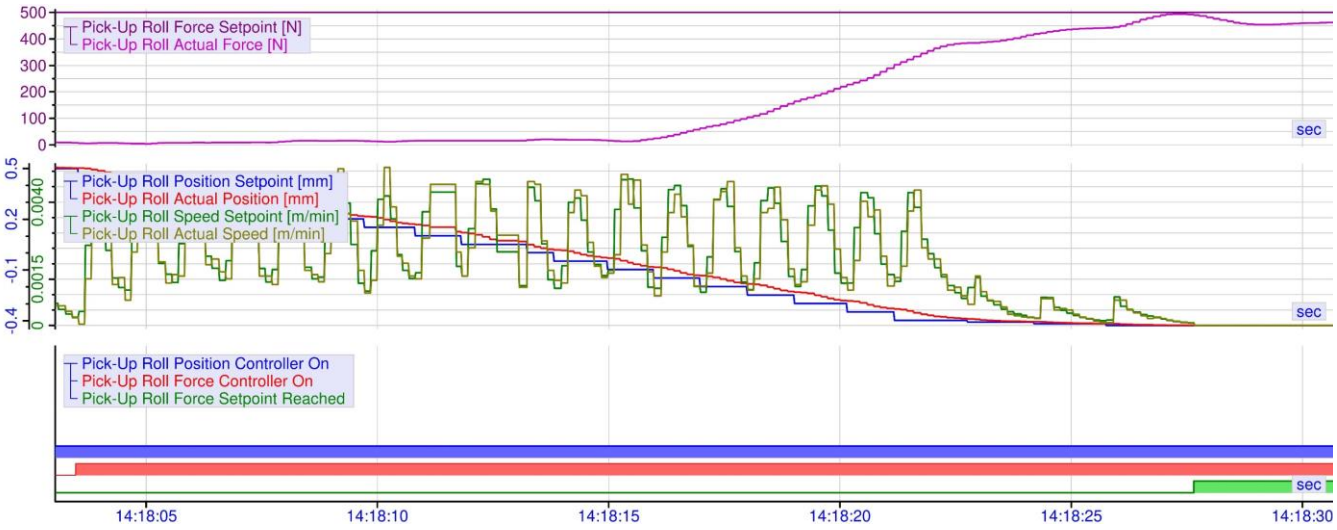


Fig. 3. Experimental verification of combined stepper-based force and position controller

A Comparison of Machine Learning Methods for Network Traffic Classification

[1]*Martin HAVRILLA (3rd year),*
Supervisor: [2]*Martin CHOVANEC*

[1,2]Dept. of computers and informatics, FEI TU of Košice, Slovak Republic

[1]martin.havrilla@tuke.sk, [2]martin.chovanec@tuke.sk

***Abstract*—The work deals with methods of analysis and characterization of network traffic using machine learning. In the first part of the thesis, selected machine learning algorithms are analyzed. Subsequently, the analysis methodology as well as the results and comparisons between individual algorithms are briefly described.**

***Keywords*—Traffic classification, Network flows, Computer learning**

I. Introduction

Nowadays, the area of network services forms a large part of the working and personal life of every person. Among the most common ways of using these services are work reasons, entertainment, shopping, etc. Therefore, it is evident that it is necessary to pay attention to security, but also to the quality of network services.

There are many methods for analyzing and characterizing computer networks such as analysis based on network flows, or direct packet analysis. Nowadays, more and more emphasis is placed on the anonymization of network traffic in order to secure it. Therefore, the analysis of network flows appears to be the most suitable way of analyzing and characterizing network traffic. For these purposes, it is possible to choose different machine learning methods, which are described together with the results of the analysis in this work.

II. The used algorithms

Due to the fact that there is a lot of irrelevant data in the investigated network, it is necessary to choose a suitable mechanism for classifying network traffic. This need is justified by the fact that the amount of inappropriate or irrelevant data introduces a number of inaccuracies into the classification and also increases the time required for the classification of computer networks.

From the point of view of machine learning necessary for the classification of network traffic, the following three algorithms were chosen, namely Naive Bayes, Decision Tree and Random Forests. These algorithms were chosen because of the relatively high accuracy in the field of data classification and also because of the relatively simple implementation of the algorithms on the data set.

A. Naive Bayes

The use of this classification method can be described as a statistical approach that is based on the probability distribution of a random event. It is based on determining the probability based on the previous state in the case that all parameters are equally valuable and independent. Then the probability for an instance of X can be determined as [1]:

$$C = \underset{j}{argmax}(p(c_j \mid X)) = \underset{c_j \epsilon C}{argmax} \frac{p(X \mid c_j)p(c_j)}{p(X)} = \underset{c_j \epsilon C}{argmax}(p(c_j) \prod_i p(x_i \mid c_j)) \quad (1)$$

While the data sample X and the dataset S is defined as n features in the dataset S. Among the main advantages of using this method of machine learning is the ease of operating with complex data sets, provided that all features are independent [2]. Based on this statement, it can be assumed that the classification of network traffic will be less accurate for parameters with variable size with respect to other network traffic parameters[3].

B. Decision Tree

It belongs to the category of divide and conquer algorithms, which are used to generate classification trees[1][3]. The main goal of this algorithm is to find the maximum profit ratio as a selection criterion of a large number of results. It is defined as[1]:

$$(S, A) = \frac{gain(S, A)}{divide(S, A)} \quad (2)$$

and individual parameters *split* and *gain* are defined as[4]:

$$split(S, A) = \sum_{i=1}^{n} \frac{|S_i|}{|S|} log_2 \frac{|S_i|}{|S|} \quad (3)$$

$$gain(S, A) = E(S) - \sum_{i=1}^{n} \frac{|S_i|}{|S|} E(S_i) \quad (4)$$

Where *E* is the entropy of *S*. Parameter S is a dataset, A is a categorical attribute. The parameter *gain* represents the information gain rate and *split* represents the split of S based on the categorical attribute A [4].

C. Random Forests

The Decision Tree algorithm is based on combinations of tree predictors[5]. Each of the trees is directly dependent on the parameters of the random vector. This vector is sampled independently and has the same distribution for all available trees in the forest. This algorithm can converge the generalization errors quite well. It is defined as a forest created with tree-structured classifiers as follows[4]:

$$h(x., \Theta_k), k = 1 \tag{5}$$

while Θ_k are random independent vectors of individual trees.

III. Analysis procedure and experimental results

First of all, it is necessary to define the basic inputs and thus the dataset itself. The dataset consists of 1,285,604 network flow records. Network traffic consisted primarily of DNS, HTTP, HTTPS, SMTP services. Subsequently, data sets containing network communication only with the mentioned network services were created. These datasets were subsequently used as a reference to the source dataset and served as a reference to look for similarities in real network traffic. Subsequent classification was carried out on the basis of parameters:

1) *Accuracy* - percentage defines the ratio of correctly classified samples to all samples. A is defined as follows[3]: $Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$
2) *Training time* - is the total training time of the examined data sample.

The variables contained in the classification parameter *Accuracy* can be defined as [3]:

1) *TP* - Successful parameter prediction A
2) *TN* - Successful parameter prediction $A^{'}$
3) *FP* - Unsuccessful parameter prediction A
4) *FN* - Unsuccessful parameter prediction $A^{'}$

A. System infrastructure

The analysis was performed on servers with OS Debian 11 with Intel Xeon Gold 5218 processor - 2.3Ghz, 16GB RAM, 750GB HDD. For the analysis itself, Weka ToolBox[6] was chosen because of its significant data mining capabilities.

B. Experimental results

At this point, the analysis was focused on obtaining the most efficient algorithm for service detection in network traffic. The main orientation was for HTTPS, SMTP, DNS services.

TABLE I
DATA CLASSIFICATION RESULTS

Parameters	Naive Bayes	Decision Tree	Random Forests
Accuracy DNS	64,57	92,37	89,24
Training Time DNS (m)	21	26	37
Accuracy HTTPS	63,29	92,44	90,07
Training Time HTTPS (m)	21	27	37
Accuracy SMTP	66,70	90,85	84,31
Training Time SMTP (m)	24	27	35

From Table I it is significant that among the three selected algorithms, the best results were achieved using the Decision Tree algorithm. In the case of the DNS Service, the accuracy of the characterization was 92.37%, for the HTTPS service 92.44% and for mail services 90.85%. The Random Forests algorithm also has comparable results. The Naive Bayes algorithm had considerably low accuracy. This finding was argued in the theoretical section above and confirmed.

The visual form of the accuracy of the characterization of network flows in the studied data set is shown in Fig.1.

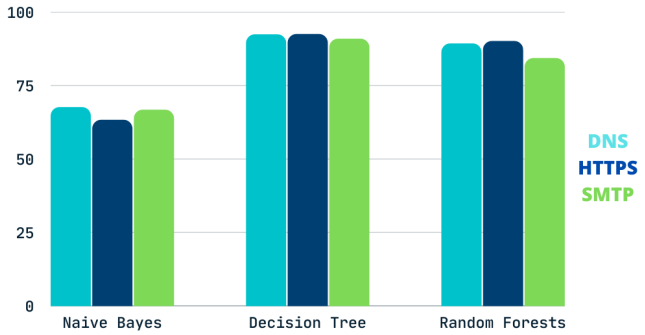


Fig. 1. Visualization of algorithm accuracy

From the point of view of data processing, it can be seen that the training time was quite long. This fact can be attributed to a relatively large source data set, as well as to a smaller reference set for individual services. From the point of view of the speed and partial efficiency of the algorithms, it could be said that the Naive Bayes algorithm is the fastest, but it does not meet the accuracy requirements compared to the other algorithms. For this reason, the Decision Tree appeared to be the most inefficient algorithm in terms of accuracy and speed.

An increase in training accuracy could occur if we filter out the trained data sample from other less important data transmission. In the same way, it would be possible to increase the processing speed.

Reclassification of computer networks can introduce inaccuracies such as loss of information, bias, lack of context, and inconsistent criteria. It is important to carefully consider potential consequences before reclassifying a network to avoid incorrect classification and misinterpretation of network characteristics.

IV. Conclusion and future work

The work created an overview of machine learning methods for characterizing some selected network services. The services were selected as the most used in network traffic as well as to find out if there is a direct difference in the classifications based on the services. It is clear from the results that the nature of the services has no direct effect on the accuracy and performance of the chosen classification algorithms. The dataset was created from network flow records from real network traffic.

Among the selected algorithms, the Decision Tree algorithm appeared to be the best, which showed the highest accuracy in the characterization of all the selected services. The claim that the Naive Bayes algorithm is not directly suitable for these purposes was also confirmed because it cannot work well with

the variable size of parameters that are often found in network traffic.

For future work, it would be appropriate to extend the classification to other used services such as e.g. FTP, VoIP, etc. It would also be necessary to classify other parameters such as Precision, False positive rate, Detection rate, etc [7]. It could be interesting to use these algorithms to characterize various forms of anomalies in network operations.

Among the last goals of future work, it would be appropriate to use these methods for the characterization of live networks in real operation.

ACKNOWLEDGMENT

This publication has been published with the support of the Operational Program Integrated Infrastructure within project: Research in the SANET Network and Possibilities of Its Further Use and Development (ITMS code: 313011W988), co-financed by the ERDF. Development and innovation of TUKE simulation environment in the field of technical sciences (004TUKE-2-1 / 2021)

REFERENCES

[1] K. Singh and S. Agrawal, "Comparative analysis of five machine learning algorithms for ip traffic classification," in *2011 International conference on emerging trends in networks and computer communications (ETNCC)*. IEEE, 2011, pp. 33–38.

[2] S. Suthaharan, "Big data classification: Problems and challenges in network intrusion prediction with machine learning," *ACM SIGMETRICS Performance Evaluation Review*, vol. 41, no. 4, pp. 70–73, 2014.

[3] J. Bakker, B. Ng, W. K. Seah, and A. Pekar, "Traffic classification with machine learning in a live network," in *2019 IFIP/IEEE Symposium on Integrated Network and Service Management (IM)*. IEEE, 2019, pp. 488–493.

[4] L. Jun, Z. Shunyi, L. Yanqing, and Z. Zailong, "Internet traffic classification using machine learning," in *2007 Second International Conference on Communications and Networking in China*. IEEE, 2007, pp. 239–243.

[5] P. Amaral, J. Dinis, P. Pinto, L. Bernardo, J. Tavares, and H. S. Mamede, "Machine learning in software defined networks: Data collection and traffic classification," in *2016 IEEE 24th International conference on network protocols (ICNP)*. IEEE, 2016, pp. 1–5.

[6] M. A. H. Morgan Kaufmann, Eibe Frank and I. H. Witten, "Traffic classification with machine learning in a live network," in *The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques"*, vol. Fourth Edition, 2016.

[7] M. Junker, R. Hoch, and A. Dengel, "On the evaluation of document analysis components by recall, precision, and accuracy," in *Proceedings of the Fifth International Conference on Document Analysis and Recognition. ICDAR'99 (Cat. No. PR00318)*. IEEE, 1999, pp. 713–716.

3D Scanning of the Indoor Environment using a quadrotor drone

[1]*Stanislav ALEXOVIČ (3rd year)*
Supervisor: [2]Milan LACKO

[1,2]Department of Electrical Engineering and Mechatronics, FEI TU of Košice, Slovak Republic

[1]stanislav.alexovic@tuke.sk, [2]milan.lacko@tuke.sk

***Abstract* —This paper presents the design and constituent elements of a quadrotor drone for the purpose of 3D scanning in indoor environments. First, the hardware components, including the drone platform, processing units, and sensors, are analysed. In addition, the software utilised for processing depth information from the depth camera is discussed. Finally, the paper highlights some challenges encountered during the development process.**

***Keywords*—3D scanning, visual SLAM, drone**

I. Introduction

3D scanning of indoor environments is the process of using 3D scanners to capture and create a digital, three-dimensional model of interior space. It involves capturing data points from surfaces or objects within the environment to generate accurate measurements and representations of the area. This technology is used for various purposes, such as architectural planning, virtual reality simulations, facility management, energy audits, and more.

Several types of sensors are used for 3D scanning, including stereo vision cameras, laser scanners, structured light scanners and time-of-flight cameras. The principle of a stereo vision camera is to capture two images from different perspectives and then use computer algorithms to combine the two images into a single 3D image[1]. Laser scanning utilises a rotating mirror to project a laser beam onto the subject while capturing its reflection with an optical detector[2]. Structured light systems employ projected patterns captured by cameras to create a 3D map from which depth information can be extracted[3]. Finally, Time-of-flight cameras measure how long it takes for emitted pulses of light to reflect off the target object before returning back to the camera[4].

This paper is based on previous work on 3D scanning an indoor environment using a hand-held 3D camera[5]. The purpose of previous work was to prove the correctness of selected sensors. In comparison, the current paper is focuses on 3D scanning using the same sensors attached to a quadrotor drone.

II. Hardware

The main hardware component is a quadrotor drone which holds all other components and enables them to fly. As the base was used a Holybro X500V2 kit. The kit includes four brushless motors, four ESCs (electronic speed controllers), PDB (power distribution board) and other propellers and mounting hardware accessories. This kit gives the user everything needed to assemble a complete quadcopter drone (see Fig. 1).

Fig. 1. Fully assembled drone with cameras and control units.

The logic of a drone sits in two control units. The first is The PixHawk 6x flying controller, which controls the drone's flight path, navigation, and stability. It receives input from the pilot or autopilot system to control the motors and other components of the drone. The controller also processes data from GPS, gyroscopes, accelerometers, magnetometers, barometric pressure sensors and ultrasonic rangefinders to maintain a stable flight.

The second control unit is the mission computer. The Raspberry Pi 4B, a single-board computer, was used for this purpose. It offers significantly improved performance over its predecessor, with a 1.5GHz 64-bit quad-core ARM Cortex-A72 CPU, 8GB of RAM, dual HDMI outputs, USB 3.0 ports and a Gigabit Ethernet port. In addition, it features 802.11ac Wi-Fi and Bluetooth 5.0 connectivity for enhanced networking. In this setup, the mission computer is responsible for processing data from the depth camera and tracking camera and providing it to the user and flying controller. The last two essential components are the primary sensors: the dept camera and the tracking camera.

The Intel RealSense Depth Camera D435i is designed for robotics, drones, AR/VR, and other applications. It features two 1280x800 resolution infrared cameras that work together to provide accurate 3D point clouds at up to 90 frames per second. The unit also includes an with three-axis acceleration and angular rate sensors, which can be used for motion tracking or pose estimation. Additionally, the camera has an RGB colour sensor with a global shutter allowing it to capture still images or video in full HD 1080p30 resolution. These features make it ideal for capturing data from dynamic

Fig. 2 From left to right. Flying a drone in the room. Output from a depth camera. Captured point cloud from the room.

environments such as autonomous vehicles.

The Intel RealSense Tracking Camera T265 provides visual SLAM to deliver accurate and reliable 6-DOF tracking. This camera utilises two image sensors, an IMU (inertial measurement unit), and proprietary algorithms to track real-time movement. In addition, the camera features a wide FOV (field of view) of up to 180° horizontal x 100° vertical and supports multiple resolutions of up to 1920x1080 pixels at 30 frames per second.

III. Software

There are two control units which need to be provided with custom software. First, the flying controller runs PX4 firmware with a slightly modified code base to enable communication over the ethernet port to connect to the ROS2 framework directly. Second, the mission computer runs Ubuntu 20.04 LTS with the ROS2 framework.

All 3D scanning is done in the RtabMap[6] ROS2 node, which needs to be fed appropriately with odometry data and the current depth map frame. Odometry is received from the tracking camera mounted on the bottom of the drone facing downward. Odometry is used for the RtabMap node and for broadcasting TF between the *odom* frame and the *base_link* frame. However, the pose is received relative to the tracking camera frame; therefore, it needs to be transformed into a *base_link* frame.

The depth map is received from the depth camera within the camera frame. However, in this case, transforming the received image into a *base_link* is easy because the TF tree is already complete, and the ROS2 framework can handle this.

Another essential software part is related to communication with the user. For this purpose, a Wi-Fi network is used, which needs to be connected to the local network. Moreover, the user's computer must also be connected to the same network to connect to the mission computer. Therefore, the user can connect directly to the mission computer and perform various tasks.

The ideal approach would be if a user could receive real-time captured point cloud of an environment. However Wi-Fi connection is not suitable for such an amount of data. Mainly, if a drone flew into an area with low Wi-Fi coverage, therefore, another approach is used. Depth images and colour images from the depth camera are streamed with the help of the GStreamer RTSP server. This method means the user can receive a video stream directly from the camera and watch it, e.g. with a VLC player. Moreover, hardware acceleration is used for encoding into h264 format, significantly reducing the mission computer CPU load. The captured point cloud is also available on the mission computer but needs to be manually downloaded.

IV. Conclusion

This paper describes the quadrotor drone employed for the purpose of 3D reconstruction of indoor environments. The drone is equipped with a depth camera that captures depth information and a tracking camera that provides the position of the drone. The paper describes the software and hardware components used to build such a drone. While almost all components were ready to use by themselves, their integration posed significant difficulties.

The main issue was a bug related to the Raspberry Pi USB port that caused the camera to stop sending its pose. Moreover, the camera manufacturer decided to abandon this product, leaving no available solution to this issue. However, the author was able to mitigate this problem by replacing the RealSense-ROS wrapper with custom code that directly interacts with the camera. This custom code was believed to be more effective in consuming tracking pose frames, thus avoiding the overflow of internal queues.

The implementation of a custom wrapper for RealSense cameras also had a positive impact on video streaming. Processing frames directly from the camera was found to be more efficient than taking them from the ROS2 topic.

The assembled drone successfully demonstrated the feasibility of 3D scanning in indoor environments and provided a robust platform for future development (see Fig. 2), including autonomous navigation without the use of GPS, collision avoidance mechanisms, and full autonomous exploration of unknown spaces.

Acknowledgements

This research paper was supported by Slovak Research and Development Agency under the APVV-18-0436 project.

References

[1] C. Lee, H. Song, B. Choi and Y. -S. Ho, "3D scene capturing using stereoscopic cameras and a time-of-flight camera," in IEEE Transactions on Consumer Electronics, vol. 57, no. 3, pp. 1370-1376, August 2011, doi: 10.1109/TCE.2011.6018896.

[2] J. P. Queralta et al., "FPGA-based Architecture for a Low-Cost 3D Lidar Design and Implementation from Multiple Rotating 2D Lidars with ROS," 2019 IEEE SENSORS, Montreal, QC, Canada, 2019, pp. 1-4, doi: 10.1109/SENSORS43011.2019.8956928.

[3] M. Atif and S. Lee, "Adaptive Pattern Resolution for Structured Light 3D Camera System," 2018 IEEE SENSORS, 2018, pp. 1-4, doi: 10.1109/ICSENS.2018.8589640.

[4] O. Choi and S. Lee, "Wide range stereo time-of-flight camera," 2012 19th IEEE International Conference on Image Processing, 2012, pp. 557-560, doi: 10.1109/ICIP.2012.6466920.

[5] S. Alexovič, „3D Scanning of the Indoor Environment" 22nd Scientific Conference of Young Researchers Proceedings, Košice, Slovakia, pp. 98-99, ISBN 978-80-553-4061-6

[6] M. Labbé and F. Michaud, "RTAB-Map as an Open-Source Lidar and Visual SLAM Library for Large-Scale and Long-Term Online Operation," in Journal of Field Robotics, vol. 36, no. 2, pp. 416–446, 2019. (Wiley)

Study of structure and molecular mobility of thermoplastic starch-based nanocomposites using NMR

[1]*Simona SAPAROVÁ (1st year)*
Supervisor: [2]Mária KOVAĽAKOVÁ

[1,2]Department of Physics, FEI TU of Košice, Slovak Republic

[1]simona.saparova@tuke.sk, [2]maria.kovalakova@tuke.sk

***Abstract*—Motivation for the study of thermoplastic starch-based nanocomposites is the need to replace conventional plastics by fully biodegradable materials due to increasing pollution of our planet by plastic waste. Starch is cheap, renewable and biodegradable polymer, which after plasticizing possesses properties typical for plastics. Its properties can be further modified by adding small amounts of nanofillers, very often clays, giving rise to thermoplastic starch-based nanocomposites. An important tool for characterization of these materials is solid-state nuclear magnetic resonance spectroscopy, which provides information on their structure and molecular mobility.**

***Keywords*—clays, nanocomposites, nuclear magnetic resonance, thermoplastic starch**

I. Introduction

Plastic materials have a large number of useful properties, such as strength, low density, formability, etc., due to which they are widely used in practice. Thus, their consumption is growing every year, not only in Europe, but also throughout the world. The negative side of this trend is that only a part of them is reused or recycled. Biodegradable polymeric materials are becoming an increasingly sought-after alternative to commonly used plastics. Bioplastics are produced by synthesis or derived from natural resources. [1] An example of a natural resource is native starch, which is very cheap and it is obtained from a wide variety of cereals, vegetables or even fruits. Plasticization of native starch gives one hundred percent biodegradable material called thermoplastic starch (TPS). TPS properties can be influenced by the type, amount and combination of plasticizers. The most commonly used plasticizers are glycerol, sorbitol, urea, formamide and other low molecular weight compounds. Another way to improve TPS properties is to add a small amount of strengthening nanofillers which results in TPS-based nanocomposite. A wide range of materials can be used as nanofillers, e.g. carbon nanotubes, graphene, cellulose, Al_2O_3, TiO_2, kaolinite, illite, montmorillonite etc. [2]

As with any newly developed material, also for biodegradable TPS-based nanocomposites it is important to know their structure and properties. Various experimental methods can be used in their study. One of them is nuclear magnetic resonance (NMR), which was discovered in the middle of the last century. Solid-state NMR has become such an important tool for polymer characterization that it is now considered a routine characterization tool. It is used not only for the study of their structure, but also molecular mobility.[3]

II. Starch and thermoplastic starch

A. *Native starch*

Starch is a natural polysaccharide that occurs in the storage organs of a large number of plants (corn, potatoes, wheat, rice etc.). Pure starch is a white, odorless, and tasteless powder. Its basic structural unit is the glucose residue $C_6H_{10}O_5$. It consists of linear amylose and branched amylopectin. Amylose makes up 17-24% of starch and the remaining 76-83% is amylopectin. The amount of amylose and amylopectin depends on the plants from which the starch was obtained. [4]

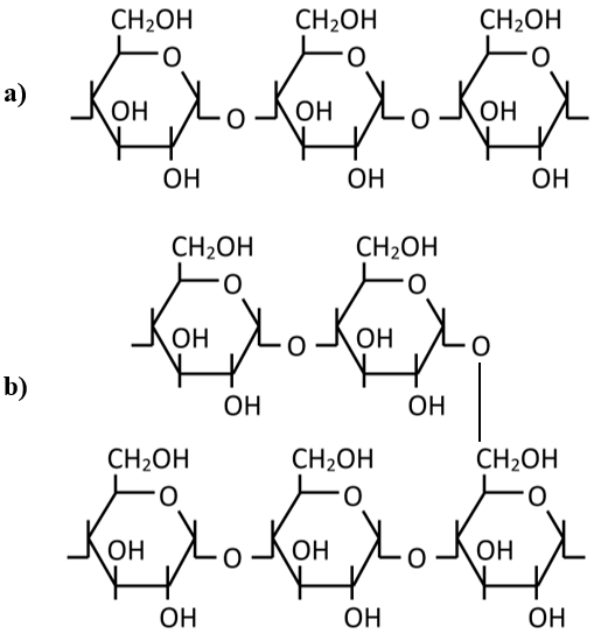


Fig. 1 Structure of amylose (a)) and amylopectin (b))

Amylose is a linear polymer whose individual monomers are connected by α-(1-4) glycosidic bonds. Amylose forms single or double left-handed helices. One turn contains 6 monomers. The primary hydroxyl groups of amylose are oriented on the outer part of the helix and cause its hydrophilicity. The secondary hydroxyl groups are oriented towards the interior of the helix and cause its hydrophobicity. Amylose is insoluble in

water. One amylose chain contains on average about 500 monomer units. [5]

Amylopectin is a branched polymer whose basic structural unit is also the glucose residue $C_6H_{10}O_5$. Individual structural units are connected by α-(1-4) glycosidic bonds. In addition to these bonds, the amylopectin macromolecule also contains α-(1-6) glycosidic bonds, by which side chains are attached to every 20-30th monomer. The degree of polymerization of amylopectin ranges from 10,000 to 100,000. Amylopectin is a semi-crystalline polymer; crystals are formed by double helices created by chain branches. It swells in water and forms colloidal solutions. [4][5]

Starch is found in plants in the form of so-called starch grains. The crystallinity of starch grains is influenced by grain size, amylose content, lipid content, degree of polymerization of the amylopectin chain, etc. Amylose and amylopectin form a polycrystalline structure in starch grains, which consists of crystalline and amorphous regions. The crystalline region is formed by parallel chains that are closely arranged next to each other. The amorphous region is composed of poorly ordered branched chains. Lack of amylose causes an increase in crystallinity. [2][5]

B. *Thermoplastic starch*

Native starch has a melting point higher than the temperature at which it degrades. For this reason, natural starch cannot be melted. The reason are the strong inter- and intramolecular starch hydrogen bonds. In order to reduce starch glass transition temperature, thermomechanical processing of starch with added plasticizers is used. The most commonly used plasticizers are water, glycerol, urea, sorbitol and formamide. In the presence of plasticizers, higher temperature (~120°C) and shear conditions, structure of native starch changes from semi-crystalline to amorphous. This process is called starch plasticization and the result is thermoplastic starch (TPS). It is a biodegradable polymer material. In general, TPS is a mixture of starch and plasticizers. The type, quantity, and combination of used plasticizers and the method of its preparation influence the properties of TPS. In the form of a gel, it has a high viscosity coefficient, in the solid phase it does not always have satisfactory mechanical properties. As the name suggests, TPS behaves like a thermoplastic, meaning it can be shaped at elevated temperatures. Another way to influence TPS properties is to create TPS-based nanocomposites. During the sample storage, starch recrystallization occurs. It is a long-term process during which the amylose molecules are preferentially arranged into a crystalline structure. TPS is widely used in various fields such as agriculture, medicine, industry, electronics, etc. [6][7]

III. NANOCOMPOSITES AND CLAYS

Composite materials consist of a matrix, which forms the continuous phase, and a filler, which is distributed in the matrix. There are several types of matrices – carbon, metal, ceramic, and polymer matrix. TPS can serve as a polymer matrix. In general, fillers can be organic or inorganic. If at least one of the dimensions of the filler is smaller than 100 nm, it is called a nanofiller. In polymer nanocomposites, four groups of materials may be used as nanofillers, namely carbon nanostructures, organic and inorganic nanofillers and clays. The last named represents the most extensive group of nanofillers. [8]

Clays can be divided into four groups (Table 1). They are characterized by a layered structure (Fig. 2). One layer consists of two types of sheets: tetrahedral and octahedral. By the arrangement of tetrahedral and octahedral sheets, clays may be specified as 1:1, 2:1, and 2:1:1 clay material. In one layer of 1:1 clay material, there is one tetrahedral and one octahedral sheet. One layer of 2:1 clay material contains two tetrahedral sheets with one octahedral sheet in between. 2:1:1 clay material has one octahedral sheet neighboring a 2:1 layer. Layer type of clay group is also listed in Tab. 1. [9]

TABLE I
GROUPS OF CLAYS AND THEIR SHEET STRUCTURE [9]

Group Name	Members	General Formula	Layer Type
Kaolinite	Kaolinite, halloysite	$Al_2Si_2O_5(OH)_4$	1:1
Smectite	Montmorillonite, talc	$(Ca,Na,H)(Al,Mg,Fe,Zn)_2(Si,Al)_4O_{10}(OH)_2$-$XH_2O$	2:1
Illite	Illite	$(K,H)Al_2(Si,Al)_4O_{10}(OH)_2$-$XH_2O$	2:1
Chlorite	Amesite, cookeite	Each member has separate formula	2:1:1

A. *Montmorillonite (MMT)*

MMT belongs to the smectite group. Its layer type is 2:1 (Fig. 2). The thickness of one layer is about 0.92 nm. Its color can be white, yellow, gray white, brownish or greenish yellow. MMT was discovered in 1847 in France, and it is a major constituent of bentonite. It has been studied for decades because of its layer expansion in the presence of water molecules. In nature, MMT is hydrophilic, and it swells in water. It has a large cation exchange and adsorption and adhesive ability, because of its large surface area. MMT attracts great attention and it is used in different areas, such as electroanalysis, drug delivery, in drilling operations and so on. MMT is commonly used as a nanofiller in polymer matrix. In polymer/clay nanocomposites, MMT can be intercalated (plasticizer molecules, polymer chains are present between the clay layers) or exfoliated (clay layers are homogenously dispersed in polymer matrix). [9][10][11]

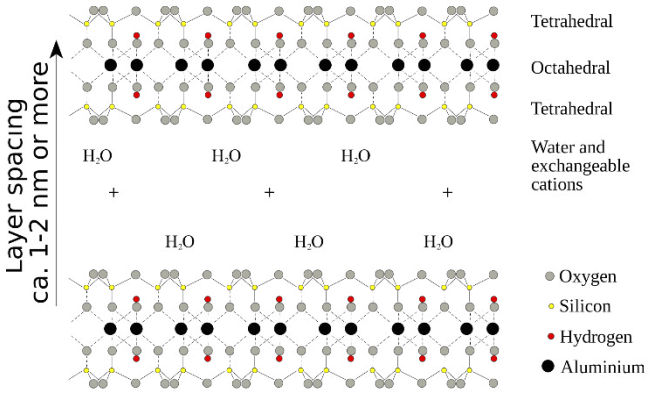


Fig. 2 Layered structure of MMT [12]

B. *Halloysite*

Chemical formula of halloysite is similar to kaolinite (Tab. 1). It has layer structure (Fig. 3) typical for kaolinite group (1:1). Halloysite is a major component of soil from volcanic materials. It can be from white to gray or brown in color. Halloysite is high biocompatible and low cytotoxic, hence, it can be used in many biological applications. It is usually multiwalled, but it can also form nanotubes. Halloysite nanotube are [15] formed due to the strain between neighboring layers of silica and alumina. The size of halloysite nanotubes varies from 20 to 4000 nm in length, from 5 to 30 nm for inner

diameter, and 20 to 200 nm for outer diameter. They are non-toxic, biocompatible and do not swell, have high porosity, high cation exchange capacity, and regeneration ability. There are two types of halloysite nanotubes – single-walled and multi-walled nanotubes. They find applications in many areas such as, drug delivery, polymer nanocomposites, electronic components, cosmetics etc. [13][14]

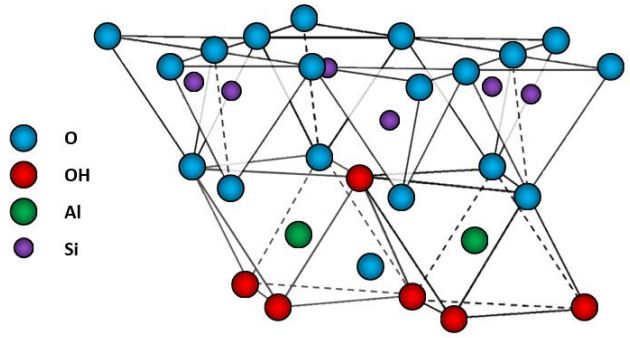


Fig. 3 Structure of halloysite (adapted from [15])

IV. NMR PHENOMENON

Nuclear magnetic resonance (NMR) is observed due to the interaction of the nuclei which possess magnetic moment with magnetic field. The magnetic moment of a nucleus is related to its spin, i.e. when a non-zero spin nucleus is placed in an external homogeneous magnetic field it interacts with this field and Zeeman splitting of energy levels occurs. If the nucleus in an external magnetic field is irradiated with a high-frequency electromagnetic field, transitions between neighboring energy levels take place. This phenomenon is called nuclear magnetic resonance. Fourier transformation of detected signal provides NMR spectrum. [3]

The shape of NMR spectrum is affected by interactions in the spin system, which include chemical shift anisotropy, dipole-dipole and quadrupole interactions. In solids, these interactions are not averaged by molecular motion and for this reason a large broadening of the signals in the spectra are observed. This drawback can be overcome by various experimental techniques. The most widely used experimental methods include magic angle spinning (MAS), heteronuclear decoupling, cross polarization (CP) and, in measurements on ^{1}H nuclei, the so-called Chen's method. [3]

Solid-state NMR spectroscopy is an important analytical tool in polymer science. Two types of characteristics can be obtained from NMR spectra. The first one is chemical shift, which is influenced by molecular structure and interactions in a sample. The second one are relaxation times, which are influenced by the chemical bonds between nuclei and local molecular motions. [16][17]

NMR spectroscopy is often used in the study of TPS-based nanocomposites. [16][17][18] In the case of study of thermoplastic starch/clay nanocomposites, measurements on ^{1}H, ^{13}C and ^{27}Al nuclei are most often carried out. ^{1}H MAS and ^{1}H BL NMR spectra provide information about molecular mobility of starch chains, plasticizers, and the effect of nanoclay on the molecular mobility. ^{13}C CP/MAS NMR is usually used to obtain information on the TPS structure. The change in amorphous/crystalline TPS phases can be evaluated by the deconvolution of C1 carbon signal in ^{13}C CP/MAS NMR spectra (Fig. 4). This enables to monitor TPS structural changes e.g. during the storage of studied samples. As far as clays as nanofillers are concerned, ^{27}Al MAS NMR are usually measured. Variations of the chemical shifts in these spectra can be a consequence of polymer intercalation. Relaxation times are often measured to monitor changes in molecular mobility being in progress in TPS-based samples. [16][17][18]

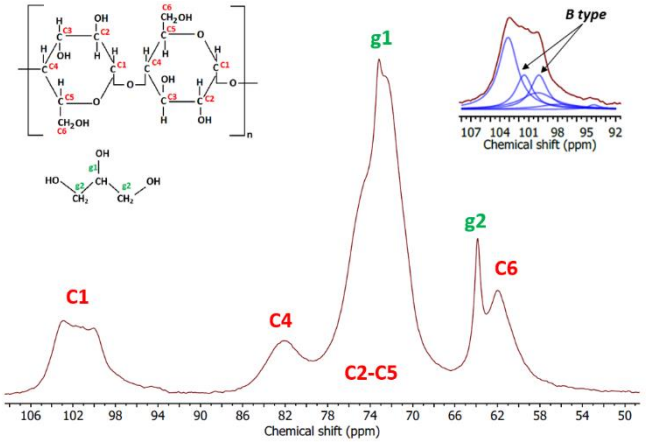


Fig. 4 ^{13}C CP/MAS NMR spectrum of starch plasticized with glycerol (adapted from [15])

V. NMR IN STUDY OF TPS/CLAY NANOCOMPOSITES

As mentioned above, the properties of TPS depend on many factors. The influence of the type of plasticizer on TPS properties is one of the most frequently studied factors. Water is the most effective plasticizer. However, water-plasticized starches exhibit poor mechanical properties. Another disadvantage is the volatility of water. Because of these reasons, besides water other plasticizers (e.g. glycerol, urea, sorbitol) are used in the preparation. The choice of secondary plasticizers depends on the potential application of the resulting material. [19]

Glycerol is one of the most studied low molecular substance as secondary plasticizer. It can have plasticizing and antiplasticizing effect on starch. With increasing content of glycerol in sample, interactions between glycerol and starch are becoming stronger. The maximum content of glycerol, when it has plasticizing effect, is 20% (w/w). In the samples with glycerol content more than 20% (w/w) but less than 50% (w/w), phase separation occurs. When the glycerol content is above 50% (w/w), starch behave like a gel. [2][19]

The other often used plasticizer is urea, which is usually used in non-food application of TPS materials. Depending on the urea content, three TPS structures are possible. When the urea content is lower than 10%, urea-starch interactions replace water-starch and inter- and intra-molecular interactions of starch. Above 10%, free urea molecules can be observed in the samples. Macroscopic phase separation occurs if the urea content is 30% or more. [20]

In many studies, the effect of glycerol, urea and mixture of glycerol and urea on structure and molecular mobility of TPS or TPS nanocomposites was investigated. The ^{1}H BL NMR spectra show that the mobility of starch chains is higher if the starch is plasticized with glycerol than if starch is plasticized with urea. From the deconvolution of ^{1}H MAS NMR spectra of starch samples plasticized with different glycerol content was observed that the water content increases with increasing glycerol content. After the sample preparation, starch is amorphous. ^{13}C CP/MAS NMR spectra of samples plasticized with glycerol and urea are different. The most significant differences are in the shape of C1 signal (Fig. 5). In the spectra of sample containing glycerol (stored for 3 months), C1 signal presents doublet typical of B-type starch structure. The shape

of this signal in the spectra of urea-plasticized sample (stored for 3 months) is typical of amorphous starch. Another factor influencing TPS structure is storage time. In the case of samples plasticized with glycerol, a decrease in the amplitude of the signal from C4 is observed during the storage. It corresponds to a decrease in the content of the amorphous phase in the sample. Changes in the shape of the signals from C2-C5 carbons and C1 are also observed due to an increase in the content of the crystalline phase. [21][22]

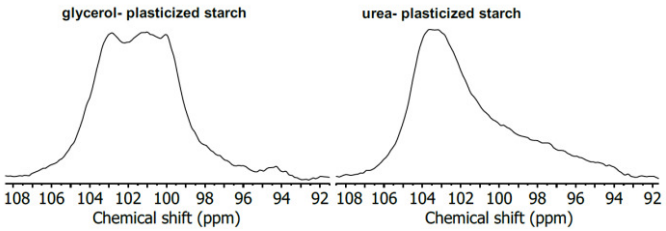


Fig. 5 Shape of C1 signal in ^{13}C CP/MAS NMR spectra of glycerol and urea-plasticized TPS/MMT nanocomposites (adapted from the thesis)

The addition of a nanofiller (for example MMT) into the TPS matrix improves its mechanical properties, since it affects the structure and molecular mobility of TPS. A comparison of ^{1}H BL NMR spectra of TPS and TPS/MMT (Fig. 6) shows that the presence of MMT causes a broadening of the signal, which is probably due to the interaction of glycerol with MMT. [23]

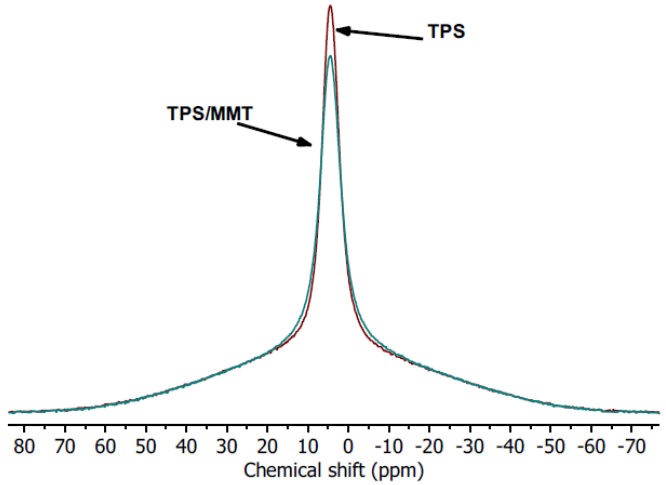


Fig. 6 ^{1}H BL NMR spectra of TPS and TPS/MMT nanocomposites (adapted from the thesis)

Besides of the effect of plasticizers and nanofiller, the influence of air humidity level on TPS and TPS/MMT nanocomposites was studied using NMR. Reduction in starch intermolecular hydrogen bond density is observed, when sample are stored at 85% relative humidity. At higher relative humidity, the higher amount of water is absorbed by the sample. It leads to an increase in the mobility of starch chains, because of additional plasticizing effect. [24]

VI. Conclusion

In recent years, research in polymer science is focused on biodegradable materials, especially biodegradable polymer nanocomposites based on thermoplastic starch. Various clays are usually used as nanofillers. These materials are studied using various experimental methods, for example scanning electron microscopy, X-ray diffraction analysis, dynamic mechanical analysis, nuclear magnetic spectroscopy, which can give important information about structure, molecular mobility and effect of many factors (storage time, type and amount of plasticizers, nanofillers etc.) on this characteristics.

Despite the extensive research so far carried out, it is still necessary to explore many possibilities for improving the mechanical properties and structure of TPS-based materials, to find suitable amounts and combinations of plasticizers, amounts and type of nanofillers, starch modifications, etc.

In the future work, TPS-based nanocomposites with various nanofillers will be studied using experimental techniques available at the Department of Physics. The other aim is to study effect of crosslinking of starch on structure and molecular mobility of nanocomposites.

References

[1] I. Vroman, L. Tighzert, "Biodegradable Polymers," in *Materials,* vol. 2, 2009, pp. 307-344.
[2] A. C. Bertolini, *STARCHES: Characterization, Properties, and Applications.* CRC Press, FL: Boca Raton, 2010.
[3] F. A. Bovey, P. A. Mirau, *NMR of Polymers.* San Diego: Academic Press, Inc. 1996, ch. 4.
[4] M. M. Burrell, "Starch: the need for improved quality or quantity-an overview," in *Journal of Experimental Botany,* vol. 54, no. 382, 2003, pp. 451-456.
[5] Y. Cornejo-Ramírez, "The structural characteristics of starches and their functional properties," in *Cyta – Journal of Food*, vol.16, no.1, 2018, pp. 1003-1017.
[6] R. Shanks, I. Kong, *Thermoplastic starch.* IntechOpen: Melbourne, 2012.
[7] A. El-Sonbati, *Thermoplastic Elastomers.* IntechOpen: Melbourne, 2012.
[8] R. Wang, S. Zheng, Y. Zheng, *Polymer Matrix Composites and Technology.* Woodhead Publishing, UK: Sawston, 2011.
[9] F. Uddin, "Clays, Nanoclays, and Montmorillonite Minerals," in *Metallurgical and Materials Transaction A.,* vol. 39, no. 12, 2008, pp. 2804-2814.
[10] M. Zoveidavianpoor, *Current Topics in the Utilization of Clay in Industrial and Medical Applications.* IntechOpen: Melbourne, 2018.
[11] S. Jayrajsinh, G. Shankar, M. Pharm, Y. K. Agrawal, L. Bakre, "Montmorillonite nanoclay as a multifaceted drug-delivery carrier: A review," in *Journal of Drug Delivery Science and Technology,* vol. 39, 2017, pp. 200-209.
[12] A. Trepte, Montmorillonite structure, available: <https://en.wikipedia.org/wiki/Montmorillonite#/media/File:Montmorillonite-en.svg >.
[13] H. A. Duarte, "Chapter 14 – Molecular Simulation of Nanosized Tubular Clay Minerals," in *Developments in Clay Science,* vol. 7, 2016, pp. 331-359.
[14] R. Kamble, M. Ghag, S. Gaikawad, B. K. Panda, "Halloysite Nanotubes and Applications: A Review," in *Journal of Advanced Scientific Research,* vol. 3, 2012, pp. 25-29.
[15] S. Saparová, "Study of morphology and molecular mobility of thermoplastic starch-based nanocomposites using NMR methods," diploma thesis, 2022.
[16] F. Zhu, "NMR spectroscopy of starch systems," in *Food Hydrocolloids,* vol. 63, 2017, pp. 611-624.
[17] P. J. Halley, L. Avérous, *Strach Polymers: From Genetic Engineering to Green Application.* Elsevier, Amsterdam, 2014.
[18] K. Jlassi, M. M. Chehimi, S. Thomas, *Clay-Polymer Nanocomposites,* Elsevier, Amsterdam, 2017.
[19] F. Ivanič, D. Jochec-Mošková, I. Janigová, I. Chodák, "Physical properties of starch plasticized by a mixture of plasticizers," in *European Polymer Journal,* vol. 93, 2017, pp. 843-849.
[20] J. Wang, F. Cheng, P. Zhu, " Structure and properties of urea-plasticized starch films with different urea contents," in *Carbohydrate Polymers*, vol. 101, 2014, pp. 1109-1115.
[21] A. Šoltýs et al, "Solid-state ^{1}H and ^{13}C NMR of corn starch plasticized with glycerol and urea," in *European Polymer Journal,* vol. 117, 2019, pp. 19-27.
[22] N. Šmídová et al., "Aging-induced structural relaxation in cornstarch plasticized with urea and glycerol," in *Journal of Applied Polymer Science,* vol. 138, 2021, 50218.
[23] A. Baran et al., "Effects of urea and glycerol mixture on morphology and molecular mobility in thermoplastic starch/montmorillonite-type nanofiller composites studied using XRD and NMR," in *Journal of Polymer Research*, vol. 29, 2022, 257.
[24] N. Šmídová et al., " Influence of Air Humidity Level on the Structure and Mechanical Properties of Thermoplastic Starch-Montmorillonite Nanocomposite during Storage," in *Materials,* vol. 16, 2023, 900.

Measuring usability of decision support system for cardiologists

[1]*Oliver LOHAJ (2nd year)*
Supervisor: [2]Ján PARALIČ

[1,2]Dept. of Cybernetics and Artificial Intelligence, FEEI TU of Košice, Slovak Republic

[1]oliver.lohaj@tuke.sk, [2]jan.paralic@tuke.sk

***Abstract*— In this article, we summarize results from usability testing of a decision support system (DSS) for cardiologists. We first proposed an extension of Van Welie layered model of usability. Based on the added aspects in extended Van Welie model, we created a questionnaire for DSS users – respondents. They followed the thought processes defined at the beginning of questionnaire and then they were asked to work on tasks within the mentioned DSS for cardiologists. After working on assigned tasks, respondents were asked to fill out the questionnaire based on their subjective perception of the system. The results from this questionnaire are provided in chapter II. and discussed in Conclusion.**

***Keywords*—decision support system, explainability, usability, Van Welie layered model of usability**

I. Introduction

Usability can be defined in many ways. The phrase "user friendliness" was replaced by the term "usability" around 30 years ago since by the early 1980s [1]. The International Organization for Standardization (ISO) has defined usability in the ISO 9241 standard series [1], which provides guidelines for evaluating and improving the usability of products, systems, and services. According to the ISO 9241 standard, usability is defined as the "extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use." The term "usability" in the context of data analysis models refers to the ease and efficacy with which users can interact with the model to obtain relevant information [4].

II. Measuring Usability

In previously published paper, we have presented an extended Van Welie model [3]. For clearer usability measurements, we divided the questions from testing usability of DSS for cardiologist into groups according to which indicator they belong to. We tested SPPRK application with students of business informatics, that took place in Scheduling and Logistics class, while we used after testing questionnaire. We mainly focused on the indicators from the newly created model published in [3].

Simplicity of control and overview of the system

Ease of use for respondents mainly depends on whether the controls and functions are easy to understand. Since the respondents were guided by the description of the steps on how to use the application to work on the result, they consider the DSS application relatively easy to use. As soon as the respondents had to work without help in the second part of the tasks (Application of hierarchical clustering), it was more difficult for them to complete the tasks. The response to the ease of working with the application would probably be lower if working with more demanding functions. On the contrary, the design of the application and clarity was comprehensible to the respondents. The overall rating of this indicator is 68.46%, which means that it needs improvement in this area. We recommend adding explanations for a better understanding of the attributes, or a more detailed description of how to work with the record filter and attribute labeling.

Explainability

When it comes to explainability, we mainly focused on whether the respondent needs a user manual, possibly the help of an expert, or whether it is necessary to already have some knowledge to work with the application. Up to 82.3% of respondents perceive that it is necessary to have instructions for using the application, while the help of a technician/expert is not necessary. The respondents stated that even the initial quick training was sufficient to understand the basic functions with which they worked. Explainability based on answers is 48.25%, which is a very low. In practice, we would recommend adding explanations to the application's functions and illustrative examples of how to work with the application. We assume that the value of this indicator would be higher if the respondents were not lay users, but experts from practice.

Accuracy of results and response rate

After examining the results of the respondents' response to this indicator, we found that it has the best rating so far - 82.93%. According to testers, the functions with which the respondents worked are well integrated. The visualization types that were applied to the application were rendered quickly and looked clear. The positive response to this question can also be attributed to the fact that respondents were explained what to look for on the graphs. The application itself does not provide a description of the graphs, which is important for users.

System functions and sustainability

Since the respondents used the basic functions they needed to work with the dataset, the answers to the question about the use of functions were mostly positive. To improve this function, it would be advisable to insert the possibility of adding a custom attribute according to the user's needs. The possibility of exporting results in several formats such as *.xlsx* or *.csv* is very useful, as the user can work with these results and format them outside the application. The application has a

well-described procedure of steps in which we should proceed to achieve the result, so the response to the logical sequence was high, 82.35%. The overall rating of this indicator is 82.63%.

Trust in the system

The questions about this indicator were directed mainly at whether the respondents find this application easier and more efficient, whether they would use it in the future or whether they would recommend it to colleagues/classmates/researchers. Trust in the application has an average value of 68.90%. We assume that this average value is mainly influenced by the fact that the respondents' frequent answer to the future use of the application was neutral. Since all the respondents when testing the application were students, this application is not so useful for them. We assume that if we conducted the testing on experts from the cardiology field, the response to the usefulness of the application would be different. This question had the highest value in this group focusing on the system trust indicator, up to 82.35%.

Trust in the result

The "result fidelity" indicator consisted of two questions that focused on whether respondents felt confident using the app. Since they only worked with a predetermined dataset, the confidence in these records was up to 88.23%. The answers to the feeling of self-confidence when using the application were different, which is also seen in the resulting value of this question - 69.41%. This value was preceded by whether the respondents had a good understanding of the procedure and the explanation of how to navigate the application. The total value of this indicator is 78.82%.

Influencing the decision

In this indicator, most of the respondents perceived that the system simplifies and facilitates their work with data, and therefore this question has an evaluation up to 78.82%. Since the respondents will not continue to work with this data and the application, the influence of the user's decision has an average value of 69.41%. The error rate was higher than we expected for some assigned tasks, which may also be because the application provides few help messages such as how to proceed and remove the error in the event of an error. The value of this indicator reached 72.54% during testing.

Preference and quality of the solution

The last indicator we focused on during testing was whether the application provides everything the user expects from it. We have seen comments that when downloading a chart as an image, the chart was not downloaded correctly. Otherwise, the application met most of the users' expectations and therefore has a response value of 77.64%. Since the application provided only minimal notifications in case of bad filtering of records or other functions, we would recommend adding error messages, why the step cannot be performed by the application and how the user can fix it. This feature would make the application faster and easier to use. This indicator acquired a value of 72.93%.

III. Conclusion

The *explainability* indicator had the smallest value. It is very important to focus on the quality of the explanation because this indicator also affects the overall impression of the application/system and also the feeling of trust in the system.

This statement is also supported by the study [4] and [5], where the authors noted that a good explanation should be understandable for people and accurately describe the model behavior in the entire feature space, which increases the value of the fidelity indicator among users. Explainability is very subjective and the perceived quality of the explanation is contextual and depends on the users, the explanation itself, as well as the type of information the interested party is interested in. In practice, we would recommend adding a user manual on how to navigate the application, illustrative examples of how to work with the application, adding explanations to the application's functions, error messages and help on how to fix them.

The indicators *system functions and sustainability* and *accuracy of results and response rate* had the highest values, where the percentage difference was minimal. Since users were working with a small amount of data, the response of the system and the generation of tables and graphs was acceptable.

The overall usability of the application reached a value of 73.61%. Based on the knowledge gained from article [6], we consider the application to be average. We also evaluated the usability of the system from the point of view of the most basic aspects such as "Effectiveness", "Efficiency" and "Satisfaction", based on the newly created usability model. We found that the indicators *explainability* and *trust in the result*, which were poorly fulfilled according to the respondents, cause low user satisfaction with the DSS. Conversely, the indicators *control, influence* and *preferences*, which belong to the aspect of *effectiveness*, had the highest values. The indicators *accuracy* and *response time* also had well filled values and therefore we consider this system effective.

The mentioned application SPPRK was tested by students at Technical university of Kosice. Because the mentioned app was developed for cardiologists, this testing with students was solely designed for approving the added aspects of usability in extended Van Welie layered model of usability. It was also interesting to see, how students with no medical background can navigate through the app and answer questions about usability of a DSS. Because the added aspects were approved by testing and proven to be useful, it opens door for extended future usability testing by cardiologists or other professional users.

Acknowledgment

This work was supported by the Slovak Research and Development Agency under contract No. APVV-17-0550 and contract No. APVV-16-0213, and by the Scientific Grant Agency of the Ministry of Education, Science, Research, and Sport of the Slovak Republic and the Slovak Academy of Sciences under grant number 1/0685/21.

References

[1] G Bevana, N., Kirakowskib, J., Maissela, Jonathan.: *"What is usability."* In: Proceedings of the 4th International Conference on HCI. 1991. p. 1-6.

[2] ISO 9241: https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en

[3] Lohaj, O.: "*Aspects of usability in clinical decision support systems*." In: 22nd Scientific Conference of Young Researchers: proceedings from the conference. Košice (Slovensko): Technická univerzita v Košiciach s. 48-51 [CD-ROM, print]. ISBN 978-80-553-4061-6

[4] Mahon, C.: "Analyzing usability testing data.", In: UX Collective, [ONLINE], Available at: < https://uxdesign.cc/analysing-usability-testing-data-97667ae4999e

[5] Zhou, J.: "Evaluating the Quality of Machine Learning Explanations: A Survey on Methods and Metrics" [ONLINE], Available at: https://www.researchgate.net/publication/349912446_Evaluating_theQuality_of_Machine_Learning_Explanations_A_Survey_on_Methods_and_Metrics

[6] Brooke, J.: *"SUS - A quick and dirty usability scale"* [ONLINE], Available at: https://www.researchgate.net/publication/228593520_SUS_A_quick_and_dirty_usability_scale

Contribution to Analysis of Thick Layers' Adhesion Fabricated by Screen Printing Technology

[1]*Peter PROVAZEK (2nd year)*
Supervisor: [2]Alena PIETRIKOVA

[1,2]Dept. of Technologies in Electronics, FEI TU of Kosice, Slovak Republic

[1]peter.provazek@tuke.sk, [2]alena.pietrikova@tuke.sk

***Abstract*—This paper presents a summary of the post-graduate study conducted over the past year, including the research tasks solved and published articles. It provides an understanding of why measuring the adhesion of printed layers on flexible substrates is essential. It also examines the factors that influence adhesion and the adhesion mechanisms most likely to be involved in the adhesion between the printed layer and the substrate. The findings from the experiments suggest that the cross-cut adhesion test can be used for measuring the adhesion of thick layers on flexible substrates fabricated by screen-printing technology.**

***Keywords*—Thick layers, adhesion measurement, cross-cut test, flexible substrates.**

I. Introduction

Flexible structures in the field of electronics are greatly impacted by printing technologies, which make possible the production of low-cost, low-weight, and low-temperature-processing structures [1]. Screen printing is the most commonly utilized method for fabricating these structures. Screen printing technology has many advantages, including the printing of conductive, semiconductive, and dielectric thick layers on flexible substrates. This widely used printing technique involves pushing a high-viscosity paste through the open area of the mask with a squeegee [2].

The quality of printed structures on a flexible substrate must be investigated, as they are subjected to mechanical stress such as twisting, bending, stretching, and deformation to different shapes. It is essential that these structures maintain stable mechanical and electrical properties during this stress [3]. The quality of printed structures can be determined by measuring the adhesion between the printed layer, with a thickness of up to 30 µm, and the polymer substrate. This adhesion helps ensure the stress resistance of the printed structure [4]. Adhesion between polymer materials is a complex topic of study involving surface chemistry, rheology of pastes, polymer chemistry, polymer physics, and stress analysis. Different adhesion mechanisms, such as those based on diffusion, mechanical, chemical, molecular, and thermodynamic principles, affect the final adhesion between foil and paste. In particular, mechanical and diffusion adhesion are the key determinants of the adhesion between the printed layer and the polymeric substrate in the context of flexible electronics [5]. Printed electronics' adhesion to flexible substrates can be affected by various factors, such as cleanliness, shape, surface roughness, and wettability. By studying the adhesion of printed layers, one can gain insight into the mechanical properties and stability of printed electronics [5]. Substrate surfaces can be modified through physical and chemical treatments to increase adhesion. This includes treatments such as plasma, flame, chemical treatments, and roughening of the substrate surface. These treatments modify the free surface energy of polymer substrates and increase the total interfacial area between the printed layer and the polymer substrate via etching [2].

II. Initial Status

Nowadays, many adhesion tests are executed with the help of professional equipment. Examples include peel tests, pull tests, and scratch tests. The main drawbacks of these methods are the length of time needed to run the tests, the cost of the required equipment, and the lack of quantitative results. Furthermore, the procedures for sample preparation and measurement are often complex [6].

The aim of my work was to investigate the current state of development of silver pastes and inks, their composition, the possibility of their printing, the flexible substrates used in flexible electronics, and the adhesion mechanism between the printed layer and the substrate. Based on this study, the topics for the thesis were chosen:

1. Development of a new polymer paste based on Ag_2O particles.
2. Analysis of layers' adhesion on flexible substrates.
3. Analysis of stable layers for flexible electrodes.

III. Solved Tasks in Previous Years

The task which is summarized in the following section were solved in the last year of postgraduate studies.

A. Thick Layers' Adhesion Measurements on Flexible Substrate by Cross Cut Adhesion Test

The cross-cut adhesion test is a useful choice for measuring adhesion in a fast and straightforward manner. Compared to other costly tests such as scratch tests, this method is reliable and easy to perform [7]. According to standards ISO 2409 [8] and ASTM D3359 [9], the cross-cut adhesion test consists of the mechanical disruption of the investigated layers using

special tools TQC CC 3000 to create six cuts perpendicular to one another, as depicted in Fig. 1. This method of human visual classification is commonly used to assess paint adhesion. This classification approach is approximate, but can be used to effectively evaluate paint adhesion. For more accurate evaluation of detached areas in fabricated electronic structures, a combination of software evaluation and an optical system should be used.

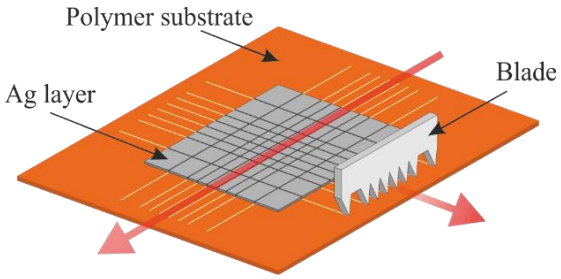


Fig. 1. Principle of the cross-cut adhesion test.

The focus of the experiment was the creation of samples using screen printing technology with three different pastes. These included two silver particle-based pastes (ED3000 and XZ250) and one carbon-based paste (ED5020). As polymer substrates, polyethylene-naphthalate (PEN) Teonex, polyethylene-terephthalate (PET) Mylar A, and polyimide (PI) Kapton HN were used.

The samples were evaluated based on the percentage of detached area from the substrate, in accordance with both the ISO and ASTM standards. The self-developed CCT Analyzer software was used to evaluate the percentage of the detached area. A score of "0" is the highest achievable score as per the ISO standard, while "5B" is the highest achievable score according to the ASTM standard [10].

TABLE I
DECISIONS LEVELS OF THE DETACHED AREAS

Classification		The size of the detached area in %
ISO	ASTM	
0	5B	0
1	4B	< 5
2	3B	< 5; 15)
3	2B	< 15; 35)
4	1B	< 35; 65)
5	0B	> 65

Table 2 displays the data from the cross-cut adhesion test method, which includes the results of the samples classified according to ISO 2409 and ASTM D3359, as well as the number of tested samples. The results of the experiment show that this method is effective for measuring the adhesion of thick layers, resulting in a classification of 66 samples into class 2 (ISO) and 3B (ASTM) on average. It appears that cross-cut adhesion testing of thick layers fabricated by the screen printing process is a cost-effective and reliable method of testing, eliminating the need for costly testing techniques. For more accurate results when testing the flexible layers, it is suitable to have a substrate thickness of at least 100 µm, to prevent over-cutting of the substrate.

TABLE 2
CLASSIFYING SAMPLES ACCORDING TO ISO AND ASTM STANDARDS

Substrate	Pastes	Samples	Detached area [%]	ISO / ASTM
PI	ED300	12	11.1	2/3B
	XZ250	12	9.7	2/3B
	ED5020	12	6	2/3B
PET	ED300	12	8.5	2/3B
	XZ250	6	9.4	2/3B
PEN	ED300	8	9.18	2/3B
	XZ250	4	12.2	2/3B

The human element is disadvantage of this method that could lead to errors such as over-cutting of the substrate, failure to make the required six cuts, or imperfect perpendicular cuts. Despite this, the results of the test are still largely unaffected, making this method suitable for testing the adhesion of thick layers.

IV. FUTURE WORK

In the future, we will focus on fabricating high-viscosity conductive paste for screen printing based on Ag_2O particles. To compare the adhesion of the self-developed paste to that of a commercially available silver paste, we will use an adhesion test method. Additionally, we will analyze the effects of surface treatments when applying these pastes.

ACKNOWLEDGMENT

This paper was supported by project KEGA No. 011TUKE-4/2023.

REFERENCES

[1] H. Ling, S. Liu, Z. Zheng, and F. Yan, 'Organic Flexible Electronics', Small Methods, vol. 2, no. 10, p. 1800070, Oct. 2018, doi: 10.1002/smtd.201800070.
[2] S. M. F. Cruz, L. A. Rocha, and J. C. Viana, 'Printing Technologies on Flexible Substrates for Printed Electronics', in Flexible Electronics, S. Rackauskas, Ed. InTech, 2018. doi: 10.5772/intechopen.76161.
[3] K. Suganuma, Introduction to Printed Electronics, vol. 74. New York, NY: Springer New York, 2014. doi: 10.1007/978-1-4614-9625-0.
[4] J. Wiklund et al., 'A Review on Printed Electronics: Fabrication Methods, Inks, Substrates, Applications and Environmental Impacts', JMMP, vol. 5, no. 3, p. 89, Aug. 2021, doi: 10.3390/jmmp5030089.
[5] A. Ghanem and Y. Lang, 'Introduction to polymer adhesion', Jan. 2017, [Online]. Available: https://www.researchgate.net/publication/319098070
[6] C. Nef, 'Analysis of Printed Electronic Adhesion, Electrical, Mechanical, and Thermal Performance for Resilient Hybrid Electronics', Dizertačná práca, University of South Florida, 2018. [Online]. Available: https://digitalcommons.usf.edu/cgi/viewcontent.cgi?article=8748&context=etd
[7] P. Provazek, A. Pietrikova, and P. Lukacs, 'Thick Layers' Adhesion Measurements on Flexible Substrate by Cross Cut Adhesion Test', in 2022 45th International Spring Seminar on Electronics Technology (ISSE), Vienna, Austria, May 2022, pp. 1–6. doi: 10.1109/ISSE54558.2022.9812806.
[8] American National Standards Institute (ANSI), ISO 2409 Paint and Varnishes Cross-cut test, New York, 2013.
[9] International, ASTM, ASTM D3359, Standard Test Methods for Measuring Adhesion by Tape Test. West Conshohocken, 2009.
[10] P. Lukacs, A. Pietrikova, and O. Kovac, 'Improvement of the evaluation of inkjet printed silver based layers' adhesion', Journal of Adhesion Science and Technology, vol. 33, no. 2, pp. 124–136, Jan. 2019, doi: 10.1080/01694243.2018.1516502.

VR systems and environments unification

[1]*Miriama MATTOVÁ (2nd year)*
Supervisor: [2]Branislav SOBOTA

[1,2]Dept. of Computer and Informatics, FEI TU of Košice, Slovak Republic

[1]miriama.mattova@tuke.sk, [2]branislav.sobota@tuke.sk

***Abstract*— This paper deals with an issue of uniform synchronization for different types of extended reality technologies. It also deals with a virtual reality and its devices, digital twins and virtual environments based on the principle of unifying diversity between such technologies.**

***Keywords*— VR, Digital twin, XR , XR Diversity**

I. Introduction

Advanced virtual reality technologies include a wide range of devices for implementing virtual spaces. Due to advanced development in a field of extended reality (XR)[1], communication and synchronization between virtual spaces and devices is much more complex [2]. The diversity is felt especially when implementing different types of virtual reality(VR) technologies like displaying devices or controlling such realities. There are many subsystems which should simulate or receive inputs from human senses such as audio systems, haptic devices or displaying devices. Therefore, it is needed to specify input and output devices to be able to implement virtual environments. Usually, tools for such development are found separately for each device. There is non-existent uniform system where it is possible to implement VR environment with a use of diverse devices without additional plugins or libraries. Also, if specific input and output devices were selected for development, environment is closely depended on these devices. Therefore, if the same environment wants to be used with another device, often it needs to be reimplemented for the needs of this device and developing environment.

II. Current State and Future Work

As it was mentioned in introduction, the implementation of VR environment usually take place in the form of choosing a specific technology for the development of a specific VR device. For example, if someone wants to create a VR environment for the Oculus Quest, it is necessary to choose a developing environment (engines, frameworks etc.) that includes development support for the glasses. The given environment, whether desktop or web-based, solves device mapping, defining semantics, transformations and instruction execution in a final VR environment. These are subsequently calculated in the final displaying device as a finished application. This means that the definition of the environment, semantics and input devices are usually handled in the hardware layer in which the system is being developed. The problem of such solution is the dependency of the support of the development environments for the given devices as well as overloading the final device with the computing processes. Based on this issue, the following requirements arose:

- Create more graphically sophisticated environments,
- be able to create more environments,
- be able to display the environment on more output devices,
- define input from controller so it can be reusable for another environment,
- have one centrum for data flow,
- specify unified semantic of inputs,
- solve unified reusable collaboration.

Those requirements can be solved with a system (figure 1.), which sins subsystem of VR environment. This subsystem should be placed on a separate computing unit which can compute graphically more sophisticated environments and have the capacity to store more of them. This subsystem should also be able to send the frames to final display devices, receive the inputs from various input devices based on the definition of the semantics from the given inputs and send outputs about the state of the environment. System should contain another subsystem for mapping various input devices and would define the semantics of their inputs as well as their values. Lastly, system would contain communication layer for data transfer.

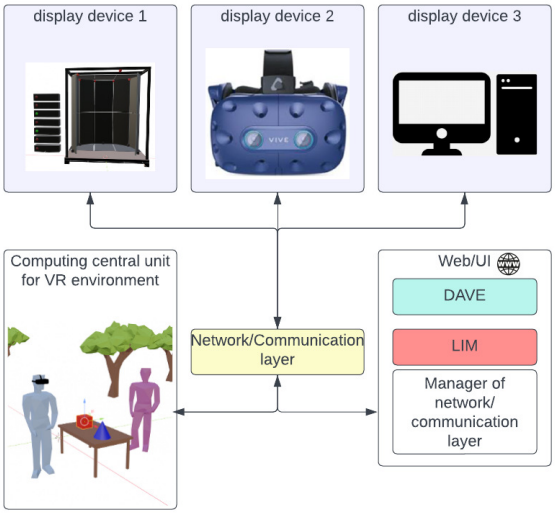


Figure 1. Simplified architecture of a system based on requirements.

Digital cave subsystem (DAVE, described in chapter V) would be the virtual helper to test the system on more complex displaying output device.

III. VR ENVIRONMENT TEMPLATE SUBSYSTEM

As was mentioned in the previous chapter, it is needed to implement one unified VR environment. Environment would contain the specially designed avatars and control scripts. Control scripts should contain the logic for sharing and management of data from avatars and objects in the scene. Number of avatars in the environments should be corresponding to the connected display devices. Definition of static, local and shared objects depend on the topic and characteristic of the environment defined by the developer. Following figure 2 represent the template of such subsystem.

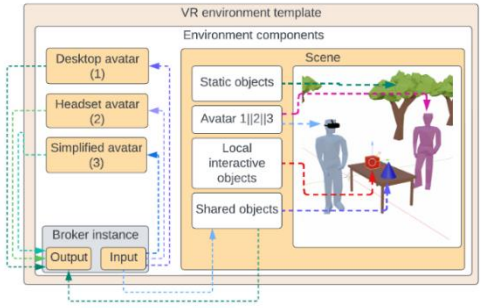


Figure 2. Template of VR environment subsystem and its components.

After display device is connected and the instance of the avatar is created, data about head, body and controller transform should be start sharing. Setting up the transformations of an avatar, shared objects, their interactions and states should come from the communication layer.

IV. INPUT SOLUTION SUBSYSTEM - LIM

To be able to control environment mentioned in a chapter before, system should contain the subsystem which can register input devices, map the values and define the data structure. Following figure 3 represents subsystem LIM (Lirkis Input Mapper)

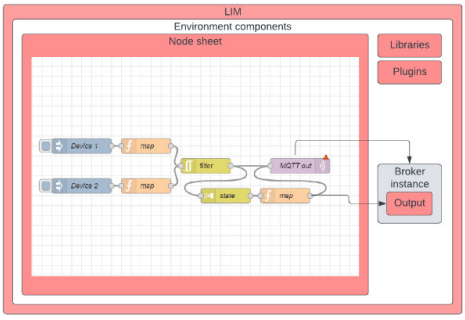


Figure 3. LIM subsystem and its components.

Node-red technology is used mostly for IOT (Internet Of Things) and is web-based. It can register local or network device and map their inputs using javascript. It contains nodes which can be connected to each other which define the final output. These outputs are then send via network layer and are linked to the specific instance of an avatar in the active VR environment.

V. DIGITAL CAVE SUBSYSTEM - DAVE

Template of VR environment subsystem should be sending the 360° frames as a visual definition from the specific avatar in the environment. As there are more complex displays then just typical output devices shaped as a rectangle, it is needed to prove the functionality of unified system on more complex final devices as well. For this issue should be created the digital twin of a LIRKIS CAVE[3] (Digital cAVE – DAVE) presented in a following figure 4.

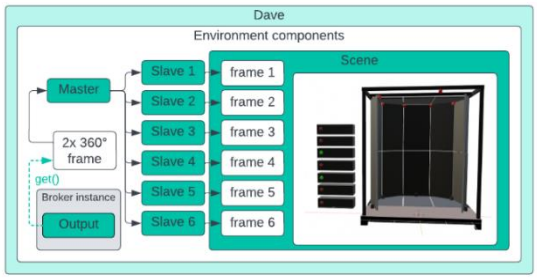


Figure 4. DAVE subsystem and its components.

Current solution of LIRKIS CAVE is not able to connect or receive data from the internet. Therefore, digital twin should be more of an improved version of the current cave solution. As the authors describe in [3] DAVE should also parallelise the frames from master to the slaves and divide the final frame into six sections. Every slave has its adjacent displays. This slave will then display the final sectioned frame to the adjacent display.

CONCLUSION

There are many issues that needs to be solved like the trustworthiness of the devices, security on the network layer, communication protocols, architecture of the avatar camera and synchronization, but there is an assumption that this system should reduce the need for creating the VR environments repeatedly, which will reduce the development time. Also, the final output device such as VR glasses would be free of computing processes as the output device will just display the frame of the remote environment. It also should be possible to display the environment on a bigger spectrum of the devices. Likewise, the need for a physical hardware to execute instructions in a virtual environment should be reduced. Ultimately, the developing environment should be unified for a larger spectrum of virtual reality technologies.

ACKNOWLEDGMENT

This work has been supported by the KEGA grant no. 048TUKE-4/2022:"Collaborative virtual reality technologies in the educational process" and the APVV grant no. APVV-21-0105 "Trustworthy human-robot and therapist-patient interaction in virtual reality"

REFERENCES

[1] Vasilchenko, A.; Li, J.; Ryskeldiev, B.; Sarcar, S.; Ochiai, Y.; Kunze, K.; Radu, I. Collaborative Learning & Co-Creation in XR. In 309 Proceedings of the Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems; Association for 310 Computing Machinery: New York, NY, USA, 2020; CHI EA '20, p. 1–4. https://doi.org/10.1145/3334480.3381056.

[2] Anthes, C., García-Hernández, R. J., Wiedemann, M. and Kranzlmüller, D. [2016]. State of the art of virtual reality technology, 2016 IEEE Aerospace Conference, IEEE, pp. 1–19.

[3] M. Hudák, Š. Korečko and B. Sobota, "On architecture and performance of LIRKIS CAVE system," 2017 8th IEEE International

Conference on Cognitive Infocommunications (CogInfoCom), 2017, pp. 000295-000300, doi: 10.1109/CogInfoCom.2017.8268260

Controllable and Photorealistic Scenes Using NeRFs

[1]*Matúš DOPIRIAK (1st year),*
Supervisor: [2]*Juraj GAZDA*

[1,2]Department of Computers and Informatics, FEI TU of Košice, Slovak Republic

[1]matus.dopiriak@tuke.sk, [2]juraj.gazda@tuke.sk

***Abstract*—Instant-NGP, a variant of neural radiance fields (NeRFs) for capturing static 3D scenes in a few minutes, is considered one of the best inventions of 2022 by the magazine Times. NeRFs belong to the category of implicit representations and parametrize the input images captured from different angles into high-fidelity 3D scenes stored in the weights of multi-layer perceptrons (MLPs). In this paper, we introduce a basic architecture of NeRFs, explain their variants towards controllable and dynamic scenes, and propose applications derived from the concepts such as metaverse, digital twin (DT) and autonomous mobility in conjunction with multi-access edge computing (MEC).**

***Keywords*—Autonomous mobility, digital twin (DT), multi-access edge computing (MEC), neural radiance field (NeRF).**

I. Introduction

The generation of photorealistic environments is extensively related to the field of computer graphics, precisely 3D rendering. These techniques construct the scene as a set of objects placed into the world coordinate system and then take into account effects such as textures, the material of the surface, visibility, shading and lighting conditions, which include reflection, refraction and diffraction. The camera viewpoint is the cornerstone for the final step within the rendering pipeline, thus placing the region of the 3D scene at a specific angle into a 2D screen. The most challenging are transparent objects and high spatial detail frequency scenes to achieve a photorealistic impression in conjunction with lighting and material.

NeRFs are eminent to become a complement to newly emerging areas such as metaverse, DT and autonomous mobility. Metaverse refers to an immersive, interactive and virtual shared space as the next generation of the Internet. The DT, defined by Glaessegen and Starge [1], is generally recognised as a complex simulation or more simulations at once to fuse physical and digital products with real-time data sharing. Autonomous vehicles tend to utilize MEC to delegate computations towards edge servers placed near to the traffic to perform high-demand computations.

II. Neural Radiance Fields (NeRFs)

NeRFs [2] belong to the family of implicit representations that utilize MLPs for the integration of a 3D static scene into the weights of the network. The purpose is to parametrize a signal as a set of images with the corresponding camera poses and overfit the MLP. The neural network produces volume density and view-dependent emitted radiance for each spatial location. A differentiable volume rendering transforms the product of MLP into a 2D image. The scene optimization is then designated by the minimization between ground truth and generated image. NeRFs accomplish a task widely known as a novel view synthesis determined for the generation of 2D images from an arbitrary angle of view.

We approximate continuous volumetric scene function with MLP F_Θ defined as:

$$F_\Theta : (\boldsymbol{x}, \boldsymbol{d}) \rightarrow (\boldsymbol{c}, \sigma), \tag{1}$$

where 5D input is composed of the location of the point $\boldsymbol{x} = (x, y, z)$ in 3D space and viewing direction $\boldsymbol{d} = (\theta, \phi)$, in which θ and ϕ represent spherical coordinates composed of the azimuthal and polar viewing angles, respectively. The output of MLP are emitted radiance $\boldsymbol{c}$ as an RGB colour and volume density $\sigma(\boldsymbol{x})$ used to determine opacity of the point at the 3D space. The network predicts the volume density $\sigma(\boldsymbol{x})$ as a function of only the location $\boldsymbol{x}$. RGB colour $\boldsymbol{c}$ is predicted as a function of both location $\boldsymbol{x}$ and viewing direction $\boldsymbol{d}$.

Rays are emitted through each pixel at a distance of specified bounds to compute the resulting colour of the pixel. In formal terms, the expected colour $C(\boldsymbol{r})$ of the camera ray $\boldsymbol{r}(t) = \boldsymbol{o} + t\boldsymbol{d}$ of its origin $\boldsymbol{o}$ and direction $\boldsymbol{d}$ in a distance of a sample t along the ray, within near and far bounds t_n and t_f, is defined as:

$$C(\boldsymbol{r}) = \int_{t_n}^{t_f} T(t)\sigma(\boldsymbol{r}(t))\boldsymbol{c}(\boldsymbol{r}(t), \boldsymbol{d})dt, \tag{2}$$

where $\boldsymbol{c}$ is emitted colour of the ray in a particular distance at a certain angle, i.e., the colour of the sample in a certain position and direction in 3D space.

Accumulated transmittance is defined as follows:

$$T(t) = exp(-\int_{t_n}^{t_f} \sigma(\boldsymbol{r}(s))ds), \tag{3}$$

where $T(t)$ is the probability that ray in the range from t_n to t_f does not hit any material.

The integral is computed numerically using stratified sampling that divides rays into N bins from which a sample was uniformly drawn. Then the numerically computed expected colour $\hat{C}(\boldsymbol{r})$ is defined as:

$$\hat{C}(\boldsymbol{r}) = \sum_{i=1}^{N} T_i(\alpha_i)\boldsymbol{c}_i, where \quad T_i = exp(-\sum_{j=1}^{i-1} \sigma_j\delta_j), \tag{4}$$

where alpha compositing $\alpha_i = 1 - exp(-\sigma_i\delta_i)$ and $\delta_i = t_{i+1} - t_i$ is the distance between consecutive samples.

A fully connected network F_Θ cannot learn high-frequency variation in colour and geometry by only feeding the model 5D input. The research of Rahaman et al. [3] proves the fact that deep neural networks need higher dimensional input to extract high-frequency variation in data. The knowledge is the base for modification of F_Θ as a composition of two functions

$F_\Theta = F'_\Theta \circ \gamma$. F'_Θ is learned by MLP and γ maps input $\mathbb{R}$ into a higher dimensional space $\mathbb{R}^{2L}$ as:

$$\gamma(p) = (sin(2^0\pi p), cos(2^0\pi p), \dots \\ \dots, sin(2^{L-1}\pi p), cos(2^{L-1}\pi p)), \tag{5}$$

where p represents each location coordinate x normalized in the range $[-1, 1]$ or three components Cartesian viewing direction unit vector $\boldsymbol{d}$ whose values lie in $[-1, 1]$.

Ray passing through each pixel leads to the redundant number of samples accumulated, especially in free space and occluded background objects. In this regard, the approach utilizes a pair of networks. A coarse network $\hat{C}_c(\boldsymbol{r})$ samples a set of N_c locations using stratified sampling defined as:

$$\hat{C}_c(\boldsymbol{r}) = \sum_{i=1}^{N_c} w_i c_i, w_i = T_i(1 - exp(-\sigma_i\delta_i)). \tag{6}$$

A fine network $\hat{C}_f(\boldsymbol{r})$ samples a set of N_f locations using inverse transform sampling. The final expected colour $\hat{C}_f(\boldsymbol{r})$ leverages all samples, thus $N_c + N_f$. Hierarchical volume sampling selects salient samples that contain parts visible in the resulting pixel.

The render function is naturally differentiable afterwards the loss is simply the total squared error between the expected colour of both coarse and fine networks and the ground truth pixel value defined as:

$$\mathcal{L} = \sum_{\boldsymbol{r}\in\mathcal{R}} [\|\hat{C}_c(\boldsymbol{r}) - C(\boldsymbol{r})\|_2^2 + \|\hat{C}_f(\boldsymbol{r}) - C(\boldsymbol{r})\|_2^2], \tag{7}$$

where $\mathcal{R}$ is the set of rays in each batch.

The fine network $\hat{C}_f(\boldsymbol{r})$ primarily determines the final rendering whereas the coarse network $\hat{C}_c(\boldsymbol{r})$ contributes to the allocation of samples in the fine network $\hat{C}_f(\boldsymbol{r})$.

Explicit representations such as meshes and voxels struggle with the size of the resulting scene. NeRFs take advantage of an efficient way to store a complex scene to a size of several megabytes. They are not dependent on dimension and resolution of the input images and are naturally differentiable as well. On the other side, training a scene composed of 100 images with a resolution 800×800 usually takes 1-2 days. Accordingly, further optimizations are necessary.

III. Extending NeRF's static scenes

Original NeRF research paper [2] has attracted the research community tremendously. In this section, we will introduce advanced approaches enhanced in inference speed, and training phase, especially extending static scenes to capture deformations and control the environment.

A. Fundamentals

Original NeRFs [2] can struggle with a generation of fine details of the 3D scene and invoke artifacts in the form of aliasing and blurring effects. If the dataset does not meet mandatory requirements, then all input images are required to be captured from the same or a similar distance. Mip-NeRF [4] extends the original NeRF to prevent this problem by taking into consideration the size and shape of the volume. The input is a 3D Gaussian approximation of conical frustum to define a region of integration of radiance field. Instead of ray casting per pixel resulting in aliasing, cone tracing is introduced. The cone is passed through each pixel and then cast into a set of conical frustums rendered towards the reduction of aliasing and blurring artifacts.

Conical frustum is defined as pentuple:

$$(\boldsymbol{o}, \boldsymbol{d}, \dot{r}, t_0, t_1), \tag{8}$$

where we emit cone from origin $\boldsymbol{o}$, thus apex of the cone, in direction $\boldsymbol{d}$. The radius is parametrized as $\dot{r}$ at the image plane defined as $\boldsymbol{o} + \boldsymbol{d}$. Authors propose integrated positional encoding (IPE) to feature a part of space in contrast to a point in space. As the result, Mip-NeRFs produce different features in the same position at different scales as opposed to the utilization of rays.

IPE is defined as follows:

$$\gamma(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \mathbb{E}_{\boldsymbol{x}\sim\mathcal{N}(\boldsymbol{\mu}_\gamma, \boldsymbol{\Sigma}_\gamma)}[\gamma(\boldsymbol{x})] = \\ = \begin{bmatrix} \sin(\boldsymbol{\mu}_\gamma) \circ exp(-(1/2)\ diag(\boldsymbol{\Sigma}_\gamma)) \\ \cos(\boldsymbol{\mu}_\gamma) \circ exp(-(1/2)\ diag(\boldsymbol{\Sigma}_\gamma)) \end{bmatrix}, \tag{9}$$

where $\boldsymbol{\mu}, \boldsymbol{\Sigma}$ represent multivariate Gaussian lifted onto the positional encoding basis with $\mathcal{N}$ levels, $\boldsymbol{x}$ is a set of positions in conical frustum and $\circ$ denotes element-wise multiplication.

Mip-NeRF reduces aliasing artifacts, significantly improves the representation of fine details and utilizes a single neural network in contrast to coarse and fine networks in the original paper. Experimental results show that the approach is 7% faster than the original NeRF with half as many parameters, and the average error rate is reduced by 17% on NeRF's single-scale dataset whereas 60% on the custom multiscale dataset.

Ref-NeRF [5] solves a problem of representing glossy or specular surfaces characterised by advanced reflections. Authors propose alternative parametrization of directional MLP to achieve more advanced interpolation of glossy appearance by providing the reflection of the viewing vector about the local normal vector. Supplementally, integrated directional encoding (IDE) is introduced to spread emittance functions between points with distinct levels of roughness. Consequently, trained scenes can be edited in material roughness. Ref-NeRF outperformed approaches including Mip-NeRF and other non-NeRF models, on the custom Shiny Blender Dataset, original NeRF dataset and Real Captured Scenes.

RawNeRF [6] is based on Mip-NeRF and omits input as processed images in low dynamic range (LDR) colour space. They lose several details in bright areas. Other details are compressed by tonemapping curve along with the following quantization to 8 bits. The proposed approach employs unprocessed high dynamic range (HDR) linear raw images to handle noise and learn more subtle details in the scene. Supplementarily, it controls viewpoint, focus, exposure and tonemapping. The approach outperforms state-of-the-art single and multi-image deep raw denoisers, mainly in the reconstruction of extremely bright and dark parts of the scene, e.g., night captures.

B. Dynamic

We have covered NeRF variants for the generation of only static scenes to query arbitrary views of the 3D space without the possibility of further interaction as motion capture.

D-NeRF [7] extends reconstruction of high-fidelity 3D scenes towards capturing rigid and non-rigid transformation over time using two MLPs. Canonical network ψ_t learns spatial mapping to encode scene, i.e., a transformation between points in time t into a canonical space as follows:

$$\psi_t : (x, y, z, t) \rightarrow (\Delta x, \Delta y, \Delta z), \tag{10}$$

where $\Delta x, \Delta y, \Delta z \in \mathbb{R}^3$ are transformed coordinates of the point in time t.

Deformation network ψ_x maps the canonical scene configuration into the deformed scene at a given time as follows:

$$\psi_x : (x + \Delta x, y + \Delta y, z + \Delta z, \theta, \phi) \rightarrow (R, G, B, \sigma). \quad (11)$$

Experimental results show that D-NeRF can perform novel view synthesis at an arbitrary time to photorealistically capture the dynamics of complex human body postures.

Hyper-NeRF [8] is an advanced NeRF variant that captures deformations of 3D scenes. The pivotal purpose is to plausibly model motion concerning topological variations in extending canonical space into higher dimensional space, called canonical hyper-space, inspired by level-set methods. Hyper-NeRF achieved high-fidelity results in the reconstruction of scenes with topological changes, e.g., a human opening their mouth in contrast to Nerfies [9] as an underlying architecture.

C. Performance

NVIDIA researchers have developed Instant neural graphics primitives (Instant-NGP) [10] primarily focused on speeding up a training phase preserving details using parallelism of graphics processing units (GPUs) by fully-fused CUDA kernels. Neural graphics primitives, mathematical functions parametrizing appearance, combined with multiresolution hash encoding reduce floating point and memory access operations. The prominent contribution concerns multiresolution hash encoding, which assumes a determined location of a point. Space is divided into grids of different resolutions representing coarse and fine features. The surrounding grids are found for the point and indices are assigned to the vertices of the grids by hashing their coordinates. Subsequently, a hash function looks up feature vectors from the corners linearly interpolated and merged into a single vector. Lastly, the outcome is fed to a shallow MLP, which trains not only their weights but encoding parameters, i.e., feature vectors. Experimental results prove a rapid reduction of training time from hours even to seconds in contrast to state-of-the-art approaches.

EfficientNeRF [11] improves the training and testing phase in conjunction with performance. Several speed optimization methods are introduced during training including lightweight MLP, valid sampling at the coarse stage, and pivotal sampling at the fine stage to efficiently master the sampling procedure. During testing, the scene is represented as tree-based data structure, called NeRFTree, to enhance rendering speed. The proposed optimization techniques reduce more than 88% of training time, and the rendering speed reaches 200 FPS.

D. Large-scale

Block-NeRF [12] elaborates large 3D scenes, especially outdoor environments, using individual Block-NeRFs and then compositing them into one unit during the rendering phase. The approach enables parallel reconstruction of parts of the scene and update regions without intervention into the rest. The proposed approach introduces appearance embeddings and learned pose refinement to tackle varying weather conditions, amount of illumination and pose artifacts in gathered dataset. Visual distinctions are adjusted by addition of camera exposure information. Transient objects, e.g., cars and pedestrians, are masked using semantic segmentation and then ignored during training. Experimental results in the form of reconstruction of the street on 2.8 million images gathered over three months prove the ability to compose large-scale environments.

Fig. 1. A PNF scene decomposed by panoptic segmentation [17].

E. Advanced control of the scene

The potential of NeRFs has attracted many researchers to push underlying static 3D scenes beyond their limits towards controlling high-fidelity and dynamic 3D scenes in time. Over the last year, many advanced research papers have been released. We primarily draw attention to editable, motion-based approaches heading towards fusion with the promising future applications mentioned in sections I and IV.

NeuralHOFusion [13] captures advanced human interactions with different types of objects including clothes or bags allowing replacements of their parts. Authors in the research paper [14] propose a controllable 3D environment allowing pose, shape and clothes control of humans based on a mesh surface. Chen et al. [15] reconstruct not only plausible looking objects but primarily interact with the environment to simulate a diverse set of deformations using point clouds. Clip-NeRF [16] controls scenes using text or image input to modify models, such as cars and chairs.

Panoptic neural fields (PNF) [17] decompose static 3D scenes into parts including background, called stuff, and countable object instances, called things, represented as instance-specific MLP in the form of object tracks. The approach addresses several tasks including semantic segmentation, instance segmentation, depth prediction, scene editing and scene decomposition into classes. Fig. 1 depicts a PNF scene decomposed by panoptic segmentation allowing extraction of objects. AutoRF [18] reconstructs a 3D scene composed of objects by decoupling the rest of the environment. Panoptic segmentation and 3D object detection extract essential data about the desired class to be reconstructed even from a single image.

IV. Applications and Future Work

As of 2023, NeRFs have gone through a major transformation from photorealistic static 3D scenes allowing the reconstruction of high-frequency reflections and transparent objects to efficiently capturing deformations through time, environmental conditions and control material, pose and shape of objects. The aforementioned core features of NeRF's variants are the source for many applications in rapidly developing areas including metaverse, DT and autonomous mobility in conjunction with MEC.

Several applications are implicitly based on the nature of the given invention. Admittedly, NeRFs can serve as a tool for human body reconstruction used in the metaverse not only for their photorealistic look but due to their ability to perform advanced interaction with the environment. For

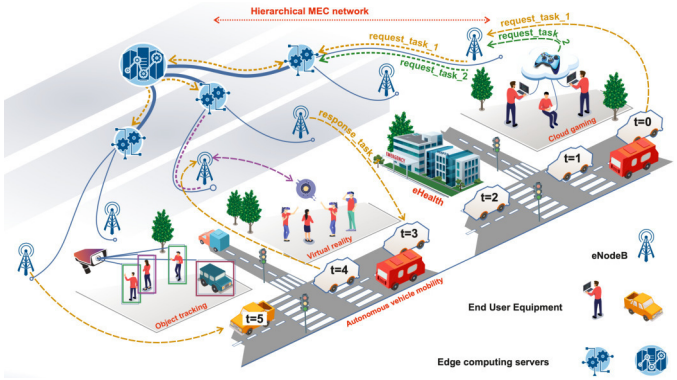


Fig. 2. MEC network applications in smart cities.

instance, model deformations are influenced by realistic material properties. The subsequent usage is connected with the development of augmented reality and the plausible placing of virtual objects into the real world along with variations in reflections. Additionally, modification of model properties, such as control of pose, shape and replacement of the parts, refines the resulting impression of mixed reality.

DT creates an immersive system by fusion of physical and digital products in real-time. It is also a part of the concept of transforming manufacturing into industry 4.0 as well. The core idea is to construct a virtual and high-fidelity form of reality via the Internet of Things (IoT), Artificial Intelligence (AI) and big data. Virtual environments do not collect only a set of heterogeneous data but replicate various processes of the physical system. Especially dynamic and controllable variants of NeRFs contribute to the prediction and visualisation of a wide range of repeating movements of robots taking place in real-time. As a result, several operations, such as advanced robot navigation, path planning and work in the manufacturing industry can be done without the intervention of a human operator represented by a specialized robot.

MEC networks can leverage base stations hierarchically arranged according to the frequency of the signal. Emitting high-frequency signals allows high speed but exhibit low stability influenced by obstacles, such as buildings, trees or even rain. Inevitably, high-frequency base stations have to be placed densely to preserve high signal quality with edge computing servers to arrange low latency and offload complex computations on end devices, e.g., cars. A principal idea of the edge servers is based on placing them near to the end devices to predominantly delegate complex computations towards the edge of the network. In this manner, it attenuates the overall cost of the end device with the absence of powerful and expensive GPUs. It signifies the bridge between processing trivial and crucial decisions with frequent offloading of the task relative to the complexity and importance. NeRFs can play an eminent role in processing of their synthetic images for computer vision tasks, such as object detection and panoptic segmentation a few seconds ahead within the predefined route of the car. Trained 3D scenes can facilitate data transfer between autonomous vehicle and edge server by sending only difference between NeRF and real image from the camera of the vehicle to reduce latency and enhance immediate interactions with the environment. Fig. 2 depicts edge servers participating in a wide range of tasks including object detection, object tracking and prompt decision-making processing of autonomous cars. The aforementioned ideas are the source for future investigations in the form of a diverse set of simulations to verify the possibilities of deployment of the proposed system into the real environment.

V. Conclusion

The research paper has introduced an underlying NeRF architecture. The cornerstone algorithm is basis for optimizations in performance, accuracy and advanced control of the 3D scene. They encompass applications in augmented reality, as avatars in the metaverse, in the reconstruction of controllable 3D scenes in conjunction with the concept of DT, and autonomous mobility using MEC networks.

Acknowledgment

This work was supported by The Slovak Research and Development Agency project no. APVV-18-0214 and no. APVV SK-CZ-RD-21-0028.

References

[1] E. Glaessgen and D. Stargel, “The digital twin paradigm for future nasa and u.s. air force vehicles,” 04 2012.
[2] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, “Nerf: Representing scenes as neural radiance fields for view synthesis,” *Communications of the ACM*, vol. 65, no. 1, pp. 99–106, 2021.
[3] N. Rahaman, A. Baratin, D. Arpit, F. Draxler, M. Lin, F. A. Hamprecht, Y. Bengio, and A. Courville, “On the spectral bias of neural networks,” 2018.
[4] J. T. Barron, B. Mildenhall, M. Tancik, P. Hedman, R. Martin-Brualla, and P. P. Srinivasan, “Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields,” *ICCV*, 2021.
[5] D. Verbin, P. Hedman, B. Mildenhall, T. E. Zickler, J. T. Barron, and P. P. Srinivasan, “Ref-nerf: Structured view-dependent appearance for neural radiance fields,” *CoRR*, vol. abs/2112.03907, 2021.
[6] B. Mildenhall, P. Hedman, R. Martin-Brualla, P. P. Srinivasan, and J. T. Barron, “Nerf in the dark: High dynamic range view synthesis from noisy raw images,” *CoRR*, vol. abs/2111.13679, 2021.
[7] A. Pumarola, E. Corona, G. Pons-Moll, and F. Moreno-Noguer, “D-nerf: Neural radiance fields for dynamic scenes,” *CoRR*, vol. abs/2011.13961, 2020.
[8] K. Park, U. Sinha, P. Hedman, J. T. Barron, S. Bouaziz, D. B. Goldman, R. Martin-Brualla, and S. M. Seitz, “Hypernerf: A higher-dimensional representation for topologically varying neural radiance fields,” *CoRR*, vol. abs/2106.13228, 2021.
[9] K. Park, U. Sinha, J. T. Barron, S. Bouaziz, D. B. Goldman, S. M. Seitz, and R. Martin-Brualla, “Nerfies: Deformable neural radiance fields,” 2020.
[10] T. Müller, A. Evans, C. Schied, and A. Keller, “Instant neural graphics primitives with a multiresolution hash encoding,” *ACM Trans. Graph.*, vol. 41, no. 4, pp. 102:1–102:15, Jul. 2022.
[11] T. Hu, S. Liu, Y. Chen, T. Shen, and J. Jia, “Efficientnerf: Efficient neural radiance fields,” 2022.
[12] M. Tancik, V. Casser, X. Yan, S. Pradhan, B. Mildenhall, P. P. Srinivasan, J. T. Barron, and H. Kretzschmar, “Block-nerf: Scalable large scene neural view synthesis,” 2022.
[13] Y. Jiang, S. Jiang, G. Sun, Z. Su, K. Guo, M. Wu, J. Yu, and L. Xu, “Neuralfusion: Neural volumetric rendering under human-object interactions,” *arXiv preprint arXiv:2202.12825*, 2022.
[14] T. Xu, Y. Fujita, and E. Matsumoto, “Surface-aligned neural radiance fields for controllable 3d human synthesis,” *CoRR*, vol. abs/2201.01683, 2022.
[15] H. Chen, E. Tretschk, T. Stuyck, P. Kadlecek, L. Kavan, E. Vouga, and C. Lassner, “Virtual elastic objects,” *CoRR*, vol. abs/2201.04623, 2022.
[16] C. Wang, M. Chai, M. He, D. Chen, and J. Liao, “Clip-nerf: Text-and-image driven manipulation of neural radiance fields,” *CoRR*, vol. abs/2112.05139, 2021.
[17] A. Kundu, K. Genova, X. Yin, A. Fathi, C. Pantofaru, L. Guibas, A. Tagliasacchi, F. Dellaert, and T. Funkhouser, “Panoptic neural fields: A semantic object-aware neural scene representation,” 2022.
[18] N. Müller, A. Simonelli, L. Porzi, S. R. Bulò, M. Nießner, and P. Kontschieder, “Autorf: Learning 3d object radiance fields from single view observations,” 2022.

Analysis and Evaluation of Data Using Artificial Intelligence for Cybersecurity

[1]*Martin HASIN (3rd year),*
Supervisor: [2]*Martin Chovanec*

[1,2]Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

[1]martin.hasin@tuke.sk, [2]martin.chovanec@tuke.sk

***Abstract*—The increasing number of cyber-attacks, such as Distributed Denial of Service attacks, is a major concern for the security of network services providing the Internet. The most common form of attack is Denial of Service attacks where attackers try to cause a website or service to fail by sending a large number of requests to the target website. To prevent such attacks, it is crucial to have a mechanism that can detect DDoS attacks in real time. Options for detecting attacks also include machine learning algorithms that show promising results in detecting Denial of Service attacks attacks. To implement these algorithms, the TensorFlow framework can be used, where data obtained from the NFStream library, which provides a description of the network using pre-defined parameters, enters the model. The most important part of training the model is the creation of a large set of records of network attacks, which are then used to create a machine learning model with the help of the interface of the tensorflow framework, which can be used to detect anomalies.**

***Keywords*—NFStream, NetFlow, Machine learning**

I. Introduction

The growth of the Internet and the increasing use of online services have resulted in an increase in cyber attacks, especially the use of Distributed Denial of Service (DDoS) attacks [1]. The HTTP protocol is used for communication between clients and servers and is an integral part of the online world that continues to expand [2]. The use of HTTP traffic for DDoS attacks is becoming more and more popular, so it is necessary to provide methods to protect against these attacks in real time. As a protection system, it is possible to use the basic technologies of HTTP protocol analysis, mainly the analysis of access to data on the server. The use of machine learning algorithms in DDoS attack detection enables real-time analysis of network traffic and has the potential to quickly and accurately detect attacks. Combining machine learning algorithms with TensorFlow and NFStream frameworks [3] can provide a robust solution for detecting DDoS attacks in HTTP traffic. TensorFlow is a machine learning and deep learning platform developed by Google [4]. The platform provides a number of tools and techniques for developing and training machine learning models, making it an ideal choice for developing a DDoS attack detection system. The NFStream platform is a high-performance and scalable network data analysis framework optimized for processing large volumes of network traffic. NFStream provides a number of tools and techniques for network traffic analysis, making it an ideal choice for HTTP analysis. In this research paper, we focus on exploring the use of machine learning algorithms with TensorFlow and NFStream frameworks to detect DDoS attacks in HTTP traffic [5].

II. Research environment architecture design

DDoS attacks are a significant threat to the availability and reliability of online services. These attacks use multiple compromised systems to overload the target network with an excessive amount of network traffic. This paper describes the design of a research environment architecture for detecting DDoS activity in HTTP traffic using machine learning algorithms. The architecture uses the TensorFlow framework and the NFStream framework to create a scalable and efficient DDoS attack detection system. The architecture is designed to handle high-volume HTTP traffic and provide fast and accurate DDoS detection at high speed or even in real time. The proposed architecture of the resulting environment consists of the following components:

- Data Collection - The first part in the architecture is to collect HTTP traffic data. Data can be collected from various sources such as routers, switches, firewalls or web server metrics. The collected data is passed through a pre-processing function where it is subsequently filtered to remove irrelevant information to reduce the size of the data and improve the processing efficiency.
- Data analyze via NFStream Framework - The NFStream library is used in the architecture as a real-time data processing and analysis tool. NFStream is a high-performance, distributed network stream processing and analysis framework optimized for real-time network traffic analysis. The NFStream library provides an API for processing network streams in real time. This library can be used to handle HTTP and HTTPS protocol usage and DDoS attack detection.
- Data Preprocessing - The preprocessed data is then passed through a series of filters to extract relevant features and remove any noise from the data. This process involves normalizing the data and transforming the data into a format suitable for use in machine learning algorithms. The resulting feature set is used as input to machine learning algorithms.
- Machine Learning Algorithms - The machine learning algorithms used in this architecture are designed to detect DDoS activity in network flows targeting the HTTP

protocol. These algorithms include supervised learning methods such as decision trees, random forests, and support vector machines, as well as unsupervised learning algorithms such as k-means clustering and self-organizing maps. Algorithms are trained on a data set containing network transmissions of the HTTP protocol as well as other protocols. Increasing the amount of data that enters the model creation process affects the accuracy, which depends on learned patterns of network traffic behavior and the behavior of network traffic associated with DDoS attacks.

- TensorFlow Framework - The TensorFlow library is used as a basic framework for creating and deploying machine learning algorithms. TensorFlow provides an API for building and training machine learning models, as well as tools for deploying the models in production environments. TensorFlow also provides support for GPU-accelerated processing, which is critical for detecting attacks from large amounts of real-time data.
- Model Deployment - Once a machine learning model is created, this model is deployed to real network traffic, focusing on the NFStream library. The deployed model is used to process incoming HTTP traffic in real time, where DDoS attacks are detected and logged.

III. Implementation of Machine Learning Model

The ability to analyze network traffic with high accuracy can make a big difference in protecting the network and its users from the disruptive and potentially devastating effects of DDoS attacks. The important step is to prepare a set of suitable data for network traffic analysis. Creating a large dataset of data is possible in several ways, using already created datasets containing various attack methods or by creating your own simulated environments for creating network attacks.

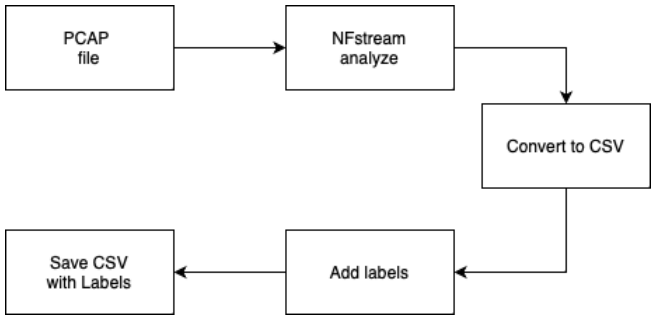


Fig. 1. Dataset analysis procedure

The fig. 1 shows the procedure for processing pre-prepared datasets, where processing results are stored and prepared data that can be implemented into a machine learning model [6]. As input data to the process, it is necessary to provide PCAP (Packet capture) samples of the required attacks on the HTTP network. These PCAPs are subsequently analyzed using the NFStream framework, and with the use of the API interface, the required parameters are converted into CSV files that can then be further processed. The most important part of the processing is the addition of LABEL parameters to the given data samples, which are additional data for the process of creating machine learning models. As a categorization for the machine learning model, the samples were labeled with numerical attributes from the number 1 to the number 5, where each number represents the type of attack in the given data sample:

- Label 1 – SYN Flood
- Label 2 – UDP Flood
- Label 3 – HTTP
- Label 4 – SCAN port
- Label 0 - Other

The HTTP type attack consisted of several attack samples and variants, specifically attacks aimed at the following vulnerabilities of the HTTP infrastructure:

- HTTP Flood (TCP, UDP)
- Slow Read
- Slow Post
- Slow Get
- HULK attack
- Goldeneye attack

These most common types of attacks were included in the datasets, which were subsequently used in the creation of a machine learning model for detecting DDoS attacks.

IV. Conclusion and Future Work

In this work, we analyzed network traffic using NFStream and subsequently created machine learning methods using TensorFlow to detect DDoS attacks in HTTP network traffic. The proposed solution used the TensorFlow library to train a deep neural network model with a dataset containing normal network traffic and network traffic containing various DDoS attacks aimed at overloading the server or network infrastructure used. Among the analyzed parts in this work was the creation of various datasets using a simulated environment or the use of currently available public datasets of various attacks on network infrastructures. The accuracy and credibility of these attacks ensures higher accuracy of the model when detecting HTTP attacks on the given network. As part of the work, a model was designed in which the metrics of the recorded network traffic are analyzed using the API interface of the NFStream framework. This framework ensures the extraction of the required parameters that describe the given network flow. With the use of the NDPI (open source deep-packet inspection) [7] library, it is possible to analyze what kind of network communication is involved, or even extract from the headers the Internet browser used. In future work, it is necessary to focus on the implantation of:

- Implement known blacklists of known IP addresses that contain unwanted network traffic into the data analysis process
- In the analysis process, implement the results from the honeypot device, which can more accurately describe possible attacks or scan the network infrastructure

By including these requirements, it is possible to increase the accuracy and detection speed of the machine learning model. The more parameters we make available to the machine learning algorithm, the more accurate a model can be trained to analyze network traffic.

Acknowledgment

This publication has been published with the support of the Operational Program Integrated Infrastructure within project: Research in the SANET Network and Possibilities of Its Further Use and Development (ITMS code: 313011W988), co-financed by the ERDF. Development and innovation of TUKE

simulation environment in the field of technical sciences (004TUKE-2-1 / 2021)

REFERENCES

[1] Y. Fu, F. Lou, F. Meng, Z. Tian, H. Zhang, and F. Jiang, "An intelligent network attack detection method based on rnn," in *2018 IEEE Third International Conference on Data Science in Cyberspace (DSC)*. IEEE, 2018, pp. 483–489.

[2] G. A. Jaafar, S. M. Abdullah, S. Ismail *et al.*, "Review of recent detection methods for http ddos attack," *Journal of Computer Networks and Communications*, vol. 2019, 2019.

[3] Z. Aouini and A. Pekar, "Nfstream: A flexible network data analysis framework," *Computer Networks*, p. 108719, 2022.

[4] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard *et al.*, "Tensorflow: a system for large-scale machine learning." in *Osdi*, vol. 16, no. 2016. Savannah, GA, USA, 2016, pp. 265–283.

[5] A. A. Galtsev and A. M. Sukhov, "Network attack detection at flow level," in *Smart Spaces and Next Generation Wired/Wireless Networking*. Springer, 2011, pp. 326–334.

[6] M. Hasin, M. Chovanec, J. Palša, and M. Havrilla, "Analysis and collection data from ip network," *Acta Electrotechnica et Informatica*, vol. 22, no. 3, pp. 18–23, 2022. [Online]. Available: https://doi.org/10.2478/aei-2022-0013

[7] G. B. Satrya, F. E. Nugroho, and T. Brotoharsono, "Improving network security-a comparison between ndpi and l7-filter," *International Journal on Information and Communication Technology (IJoICT)*, vol. 2, no. 2, pp. 11–11, 2016.

Comparison of various magnetoelastic compressive force sensor models

[1]*Šimon GANS (2nd year)*
Supervisor: [2]Ján MOLNÁR

[1,2]Dept. of Theoretical and Industrial Electrical Engineering, FEI TU of Košice, Slovak Republic

[1]simon.gans@tuke.sk, [2]jan.molnar@tuke.sk

***Abstract*—This paper addresses the problem of a meaningful comparison of various magnetoelastic sensors. Because the mechanical and electromagnetic fields inside the sensor bodies can quantitatively differ quite drastically between them and mechanical loading can be applied to various surfaces of the sensor, comparing them solely just on their electrical output signals is not enough to compare them fairly. Therefore a more intricate classification system is introduced.**

***Keywords*—magnetism, magnetoelastic, mechanical, sensors**

I. Introduction

Magnetoelastic sensors have gained a lot of popularity in the field of electrical and non-electrical measurement. They have been found useful for the sensing of compressive[1], tensile[2], bending[3], and torsional forces[4]. When manufactured into thin wires and ribbons, they can be used for temperature and pressure sensing[5], the measurement of the environment's pH[5] or viscosity[6], and much more. When coated with glass, the sensor's body is protected from harsh environments and can operate under nearly any conditions[7].

The sensing principle of such sensors is based on the change of magnetic properties due to mechanical action. A stress-induced change of permeability is evaluated, which mostly manifests itself as a change of coil inductance or a change of coupling between transformer coils[8]. Other evaluation principles like a change in mechanical resonant frequency or a change of harmonic contents in the output signal are also used[9].

In the field of compressive force measurement, they are mostly used for the measurement of very large forces that are present in rolling mills[11], or bridge structure monitoring[10]. For other everyday applications like personal or kitchen scales usually, resistive tensometric force sensors are used, due to their excellent linearity and their ability to work with DC input. On the other hand, magnetoelastic sensors made from bulk materials tend to be linear in a small interval of forces and they usually require AC waveforms to operate[8]. However, they are mechanically extremely robust, and their output signal is many times larger and is more sensitive to stress than it is in the case of tensometers. They are also immune to electromagnetic interference[8]. These factors invite us to spend resources and time optimizing such sensors to possibly broaden their application scope in today's industry.

However, when choosing a metric of how well a sensor performs, we suggest that it is not enough to just focus on the output characteristics of the sensor, but also on how the loading affects the sensor's structural integrity, the linearity of the sensor's response, and more. Therefore a new metric of quantitative sensor comparison is proposed.

II. Sensor Characteristics

Multiple sensor characteristics were chosen that characterize it individually from the mechanical and electromagnetic points of view, which will be described in this chapter. The analysis steps will be demonstrated on the shown sensor model (Fig. 1).

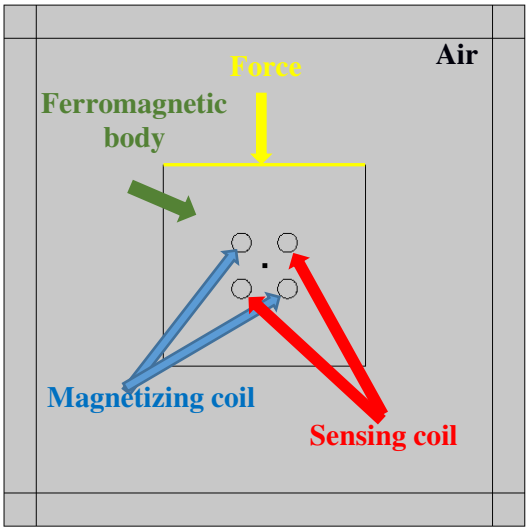


Fig. 1: An example 2D model of an elastomagnetic sensor that is the subject of analysis.

A. Output voltage

For the measurement of static compressive forces, an AC input energy source is mostly used[8]. Permanent magnet excitation has been employed as well, but it is more commonly used for dynamic force measurement[12]. The output voltage of the sensor, therefore, is a periodic signal, which characteristics vary with the action of force. The amplitude, RMS value, and frequency components vary with force application. The RMS value was chosen for the evaluation of this change (1),

$$U_{RMS} = \sqrt{\frac{1}{T}\int_{t}^{t+T} u^2(t)dt} \qquad (1)$$

where u is the output voltage waveform and T is the period of the waveform. The change of RMS voltage as a function of force is then considered. A metric that characterizes the performance of the sensor from the viewpoint of this characteristic is the sensitivity of the sensor (the slope of the line in Fig. *3*).

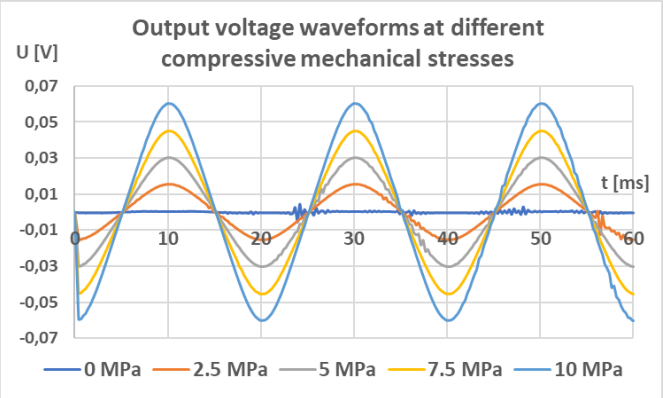


Fig. 2: The output voltage waveforms (voltages induced on the secondary coil) of the analyzed sensor at different compressive mechanical stresses. The magnetizing current flowing through the magnetizing coil was a sine wave with 1 A amplitude and 50 Hz frequency.

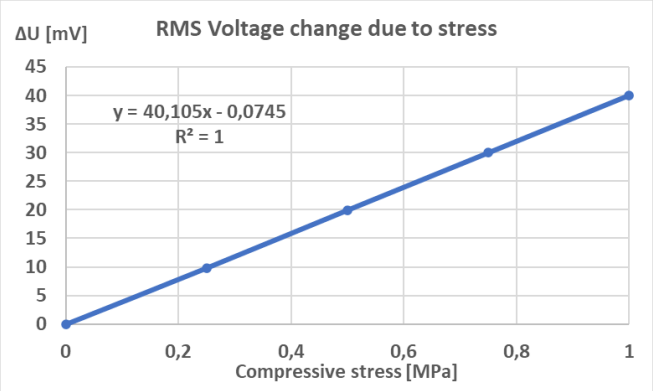


Fig. 3: Graph of the RMS voltage changes computed from the waveform in Fig. 2 as a function of compressive stress.

B. Mechanical loading

When a force is acting on mechanical structures, the analysis of internal stresses is an important branch of mechanical engineering. Every structure has a maximum yield point after which the structure mechanically fails or plastically deforms. In the case of sensor technology, measurement repeatability is an important factor so the material should be loaded only in its linear elastic region of the corresponding stress-strain curve to ensure that the internal structure does not undergo any changes. Loading the sensor further can lead to irreversible deformation or material failure which drastically changes the sensor's characteristics. To estimate the amount of stress a given sensor under load experiences, with the assumption that the material behaves like a linear elastic material, the Huber-Mises-Hencky theory of maximum energy of distortion was employed. The output of this analysis is the von Mises stress distribution inside the material[13]. The analysis was also done with the help of finite element analysis models.

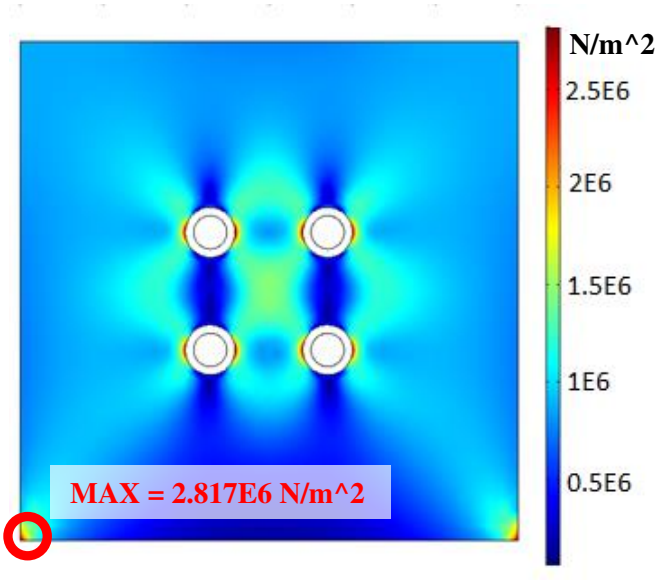


Fig. 4: 2D plot of the internal mechanical stress analysis (von Mises) and the point of maximum stress highlighted by the red circle. The input load was a mechanical compressive stress of 1 MPa in the force direction shown in Fig. 1.

The stress alone is not enough to evaluate the mechanical strength of the structure. The applied load and its placement must be considered because internal stresses vary when loads act in different regions of the body. The maximum stress observed in the stress analysis of the sensor body is divided by the mechanical stress load acting on the body (force divided by unit area). This gives a number that is invariant of load size due to the linearity of the analysis and defines the mechanical structure strength of the sensor in correspondence to the given load placement. The higher this number is, the worse the structure performs mechanically. It is described by equation (2),

$$x_{mech} = \frac{\sigma_{max}}{\sigma_{load}} \tag{2}$$

where x_{mech} is the described mechanical structure metric, σ_{max} is the mechanical stress at the most stressed point of the material and σ_{load} is the force per unit area that is being applied to the sensor body. The x_{mech} value for this sensor is 2,817.

C. Static characteristic nonlinearity

An ideal elastomagnetic sensor has a linear relationship between the output voltage change and the acting force. When non-linearities are introduced, then the force that acts on the sensor must be interpolated (or in some cases extrapolated) from a look-up table or it must be computed from some fitting formula, which makes the evaluation process more computationally expensive and prone to errors. The maximum nonlinearity error is computed in reference to the linear regression curve that is fitted from the static sensor characteristic shown in Fig. *3*[15].

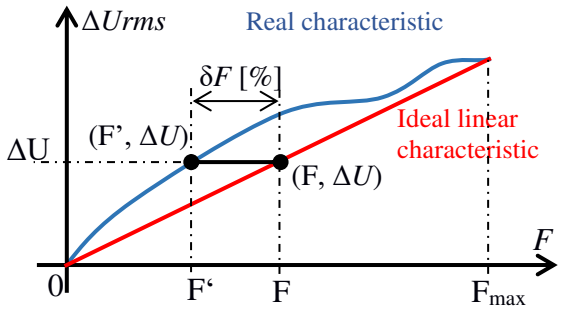


Fig. 5: An example showing how the nonlinearity error is computed.

The linear model is given by equation (3),

$$\Delta U_{RMS} = aF + b \tag{3}$$

where a and b are constants, F is the applied force to the sensor and ΔU_{RMS} is the output RMS voltage of the sensor compared to the unstressed state. The maximum relative nonlinearity error of the force measurement is defined by equation (4),

$$\delta F = \max\left(\left|\frac{aF' - \Delta U_{RMS} + b}{a \cdot F_{max}}\right| \cdot 100\right) \tag{4}$$

where F' is the corresponding force for the given RMS voltage change in the real sensor characteristic and $Fmax$ is the maximum force that was applied to the sensor.

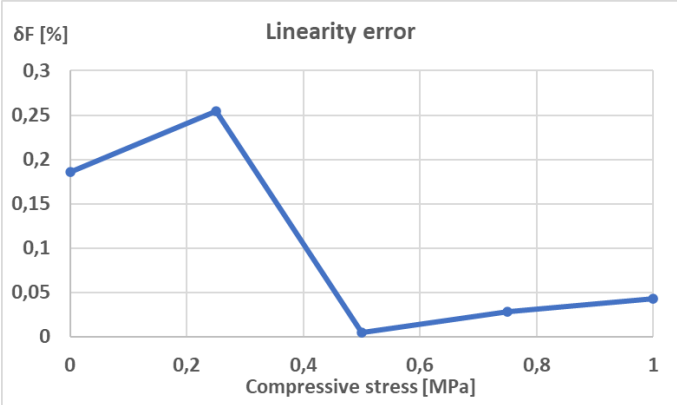


Fig. 6: The relative linearity error computed from the output voltage data of the analyzed sensor. The maximum linearity error was 0,251 %.

III. The Final Comparison Metric

Evaluating the performance of different sensors numerically should be a combination of all mentioned individual metrics. An approach was introduced in which many sensors were simulated and analyzed by the specified metrics mentioned in the last chapter. The metrics are then normalized in between the set of sensors so that every sensor gets assigned a value between 0 and 1 in the given metric linearly. The metric will then get combined into a final value that characterizes the performance of the sensor for static load measurements. This metric orders the set of sensors performance-wise. The formula for the final characteristic is expressed by (5),

$$y_{fin} = ay_{output} + by_{mech} + cy_{lin} \tag{5}$$

where y_{fin} is the final metric of the sensor performance, y_{output} is the normalized metric that quantifies the sensor performance, y_{mech} is the normalized mechanical loading characteristic of the sensor and y_{lin} is the normalized nonlinearity error characteristic and a, b, and c are constants of proportionality that determine the importance of the characteristics in the final metric. For the comparison, we used a = 2, b = -1, and c = -2, which means that we set the linearity and the sensitivity to be twice as important compared to the mechanical metric. The negative numbers were assigned to a metric in which higher numbers define a worse-performing sensor (a higher nonlinearity error is worse than a smaller the same applies to the mechanical characteristic). Some problems can arise with the usage of this type of sensor description as when the set contains a sensor that is numerically very different from the others metric-wise, then the other sensors are more closely grouped in the normalized interval which diminishes their differences. Therefore, when a sensor is very underperforming compared to the rest it should not be added to the comparison set.

IV. Results

A set of 16 sensors were simulated and analyzed by the given metric. The sensor models can be obtained from the 1st author of this work and were not included in this paper, because of length considerations. The comparison graph of the sensor performances can be seen in the graph below (Fig. 7).

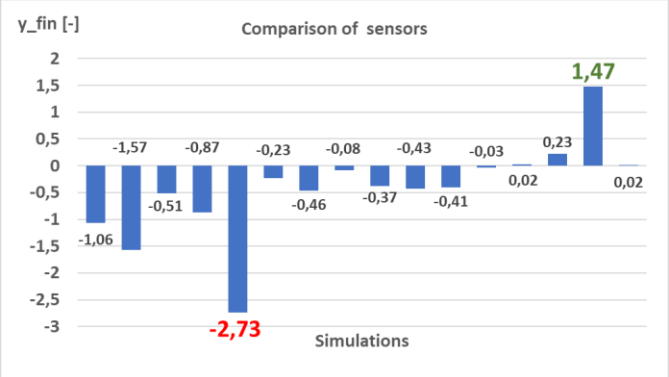


Fig. 7: A graph of different sensors' final metrics. We can see that the largest value of the metric is 1,47 and it belongs to the sensor depicted in Fig. 1. It combines high sensitivity, high linearity, and also good mechanical characteristics from the set of simulated sensors. The lowest value of -2,73 describes the performance of a sensor that is depicted in Fig. 8.

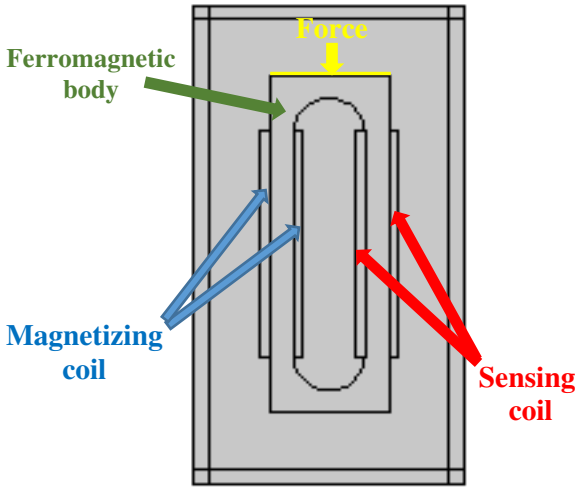


Fig. 8: The worst-performing sensor geometry with a performance metric of -2,73.

V. FUTURE WORK

This is one example of a metric that will be used as a cost function in the optimization process of the depicted sensor geometry (Fig. 1) and its characteristics to maximize its performance. Parameters like coil turn count, input current amplitude and waveform, dimensions of the sensor, placement of holes of the coil, and their shapes and orientation will be varied to find the optimum configuration for a given scenario. The output of this process should be an optimized sensor geometry with optimized input power characteristics. The analysis of such a model and its experimental verification will be the subject of future work.

An example of a geometry variation of the sensor depicted in Fig. 1 is shown in Fig. 9. The "sides" of the sensor are parametrically defined via an ellipse arch with a given eccentricity and the holes for the coils are defined by their eccentricity, orientation, and position. By sweeping through many different combinations of parameters, the best-performing sensor will be selected and tested experimentally in reference to worse-performing sensors to confirm the reliability of the model.

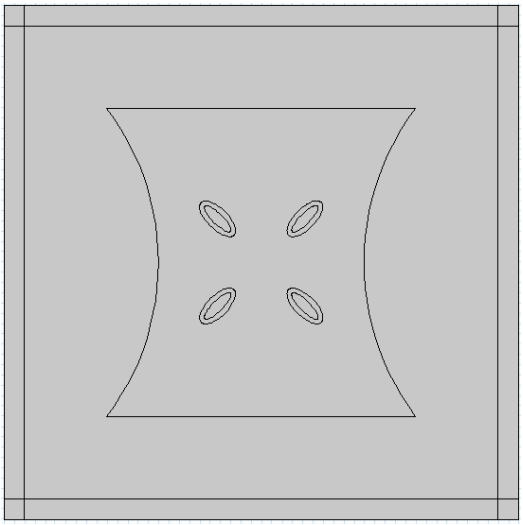


Fig. 9: Parametric model of the sensor body

ACKNOWLEDGMENT

The authors thank for the financial support from the GRANT FEI 2023 project for the grant specified by the index FEI-2023-91.

REFERENCES

[1] A. Bieńkowski, R. Szewczyk, J. Salach, Industrial Application of Magnetoelastic Force and Torque Sensors, 2010, Acta Physica Polonica Series a, 118(5):1008-1009. DOI: 10.12693/APhysPolA.118.1008. Available on the internet at: http://przyrbwn.icm.edu.pl/APP/PDF/118/a118z5p120.pdf

[2] M. Nowicki, Tensductor-Amorphous Alloy Based Magnetoelastic Tensile Force Sensor, 2018, 18(12), 4420, MDPI Sensors. DOI: 10.3390/s18124420. Available on the internet at: https://www.mdpi.com/1424-8220/18/12/4420

[3] P. Grenda, M. Kutyła, M. Nowicki, T. Charubin, Bendductor – Transformer Steel Magnetomechanical Force Sensor, 2021, MDPI Sensors, 21(24), 8250. DOI: 10.3390/s21248250. Available on the internet at: https://www.mdpi.com/1424-8220/21/24/8250

[4] A. Ostaszewska-Liżewska, M. Nowicki, R. Szewczyk, M. Malinen, A FEM-Based Optimization Method for Driving Frequency of Concatless Magnetoelastic Torque Sensors in Steel Shafts, MDPI Materials, 2021, 14(17), 4996. DOI: 10.3390/ma14174996. Available on the internet: https://www.mdpi.com/1996-1944/14/17/4996?type=check_update&version=2

[5] C. A. Grimes, S. C. Roy, S. Rani, Q. Cai, Theory Instrumentation and Application of Magnetoelastic Resonance Sensors: A Review, MDPI Sensors, 2011, 11(3):2809-2844. DOI: 10.3390/s110302809. Available on the internet: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3231618/

[6] P. Chen, Q. Jiang, S. Horikawa, S. Li, Magnetoelastic-Sensor Integrated Microfluid Chip for the Measurement of Blood Plasma Viscosity, 2017, Journal of The Electrochemical Society, 164 B247. DOI: 10.1149/2.1441706jes. Available on the internet: https://iopscience.iop.org/article/10.1149/2.1441706jes

[7] E. Komová, M. Varga, R. Varga, P. Vojtaník, J. Torrejon, M. Provencio, M. Vázquez, 2008, Acta Physica Polonica A, 113 (2008), No. 1. DOI: 10.12693/APhysPolA.113.135. Available on the internet at: http://przyrbwn.icm.edu.pl/APP/PDF/113/a113z1033.pdf

[8] O. Dahle, The Pressductor and the Torductor – Two Heavy-Duty Transducers Based on Magnetic Stress Sensitivity, 1964, IEEE Transactions on Communication and Electronics. DOI: 10.1109/TCOME.1964.6592601. Available on the internet at: https://ieeexplore.ieee.org/document/6592601/authors#authors

[9] C. A. Grimes, C. S. Mungle, K. Zeng, M. K. Jain, W. R. Dreschel, M. Paulose, K. G. Ong, Wireless Magnetoelastic Resonance Sensors: A Critical Review, 2002, MDPI Sensors, 2(7). DOI: 10.3390/s20700294. Available on the internet: https://www.mdpi.com/1424-8220/2/7/294

[10] X. Li, J. Zhou, L. Zhang, J. Yang, Y. Xiao, Automatic monitoring of continuous rigid frame bridges by a magneto-elastic effect method, 2017. ACTA Press. 2017 International Journal of Robotics and Automation. DOI: 10.2316/Journal.206.2017.1.206-4701. Available on the internet at: https://www.actapress.com/PaperInfo.aspx?paperId=45454

[11] L. Jonsson, 50 years of flatness control with ABB's Stressometer, ABB Review, 2017. Available on the internet at: https://library.e.abb.com/public/d83ad7d9b0b8438193e9fe7e67a2a03a/50%20Years%20of%20Flatness%20Control%20with%20ABB's%20Stressometer.pdf

[12] M. J. Dapino, F. T. Calkins, R. C. Smith, A. B. Flatau, A Magnetoelastic model for magnetostrictive sensors, 1994, North Carolina State University, Center for Research in Scientific Computation, Raleigh, NC, 27695-8205. Available on the internet at: https://www.researchgate.net/publication/235108750_A_Magnetoelastic_Model_For_Magnetostrictive_Sensors

[13] J. Bocko, M. Pástor, I. Delyová, M. Hagara, Elasticity and strength I., Technical University of Košice, 2021, 1st edition, pages 86-91. ISBN 978-80-553-3915-3.

[14] J. Fraden, Handbook of Modern Sensors: Physics, Designs and Applications, 5th edition

[15] R. Hogan, How to Calculate Linearity Uncertainty, 2019, ISObudgets. Available on the internet at: https://www.isobudgets.com/how-to-calculate-linearity-uncertainty/

Influence of temperature on domain wall dynamics in rapidly-changing magnetic field

[1]*Simeon SAMUHEL (3rd year)*
Supervisor: [2]*Jozef ONUFER*

[1,2]Dept. of Physics, FEI TU of Košice, Slovak Republic

[1]simeon.samuhel@tuke.sk, [2]jozef.onufer@tuke.sk

***Abstract*— This paper is a summarization of the last year of post-gradual study. Amorphous glass-coated microwires are perfect materials for theoretical study and technical applications. The study of the dynamic of a domain wall (DW) moving in bistable microwire in a region of rapidly-changing magnetic field under influence of lower temperature is presented. Peaks in wall velocity were observed as DW passed through the region of rapidly increased magnetic field. The propagating DW is forced to change its parameters in a very short time, resulting in an increase in velocity above its equilibrium value after the DW passes through a region of rapid magnetic field increase.**

***Keywords*— Glass-coated microwires, domain wall dynamics, bi-stable behavior.**

INTRODUCTION

Mechanical stress and strain sensors are increasingly used in countless industrial applications like aerospace, automotive, process control, biotechnology, etc. [1-3]. Therefore, monitoring and recording the strains is crucial for the actual research and further analysis [4,5]. Nowadays, Structural Health Monitoring is often used, which involves the observation and analysis of a structural system over time using periodically sampled response measurements to monitor changes in the material and geometric properties of bridges and buildings. Magnetic-based sensors represent the possibility for contactless mechanical strain and stress monitoring.

A specific group of magnetic materials for this application are ferromagnetic amorphous glass-covered microwires. They consist of a metal core with a diameter of a few micrometers to 100 μm covered with a glass envelope with a thickness of a few micrometers up to 25 μm. They are prepared by drawing a glass capillary with molten melt through a stream of water, the so-called Taylor-Ulitovsky method [6]. It is known that the domain structure of microwires with positive magnetostriction consists of a large axial domain in the metal core, which is surrounded by many antiparallel radial domains in the shell [7]. In addition, small closure domains (domain structures) appear at the wire ends to reduce stray fields [8, 9]. Magnetization reversal occurs with one large Barkhausen jump.

For understanding the DW dynamics in amorphous microwires, the factors limiting the DW velocity as well as searching for the ways allowing further DW velocities improvement is therefore essentially relevant for the implementation in applications involving DW propagation in microwires.

In this work, we present an experiment, which allows the dynamics of a DW propagating in a rapidly changing driving magnetic field to be studied.

I. EXPERIMENTAL

We studied the most typical Fe-rich amorphous glass-coated microwire: $Fe_{77.5}Si_{15}B_{7.5}$ (metallic core diameter d = 7.5 μm, total diameter D = 15 μm and sample length was 12.5 cm) with positive magnetostriction coefficient λ_s.

An experiment was set up to study the dynamics of one DW in bistable microwires in an inhomogeneous magnetic field. Its arrangement is schematically shown in Fig. 1. The bistable microwire is placed in two coaxial magnetizing coils So and So1. DW moves from one end of the microwire while passing through the coils So1(R=0.3 mm, l=30 mm, number of turns z=864) and through long pick-up coil PuC (l_1=20 mm, number of turns n=285). Solenoid So1 creates the magnetic field H_1 and long solenoid (So) creates the magnetic field H. PuC is used to monitor the single DW motion along the microwire. The region close to the end of solenoid So1 is the source of an inhomogeneous magnetic field. The end of solenoid So1 is located approximately in the centre of the PuC. The experimental method is described in detail in the article [9].

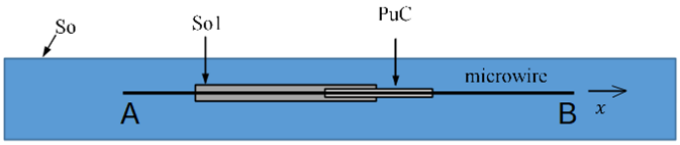


Fig. 1. Experimental set-up

For an experimental temperature measurement, the coil system (mentioned above) is placed in a Dewar vessel filled with liquid nitrogen. The sample holder with a system of coils can be moved up or down over the surface of the liquid nitrogen using a computer-controlled stepper motor to change or stabilize the temperature. A thermometer (based on a Pt100 resistor) is used to measure the temperature. Such an experimental set-up allows the measurement of velocity $v(H)$ dependence at a constant temperature.

II. RESULTS AND DISCUSSION

We wanted to find out how temperature affects the shape of the observed velocity peaks in the inhomogeneous magnetic field. Samples with strong unidirectional effects [10, 11] were chosen for our measurements.

The processing of induced voltage signals [12] at different magnetic fields H confirmed the presence of a unidirectional effect, and the measurement at different temperatures gave us local $v(H)$ dependences for fast and slow DW (Fig. 2).

The presented results of measurements on microwires with unidirectional effect show that *v(H)* dependencies of fast (closed symbol ■) and slow (open symbol □) DWs are less steep with decreasing temperature. This effect is more significant for fast DW than for slow DW.

The effect of temperature is very similar to the effect of applied tensile stress. [12, 13]. The source of tensile stress are probably different thermal expansions coefficients of metal and glass.

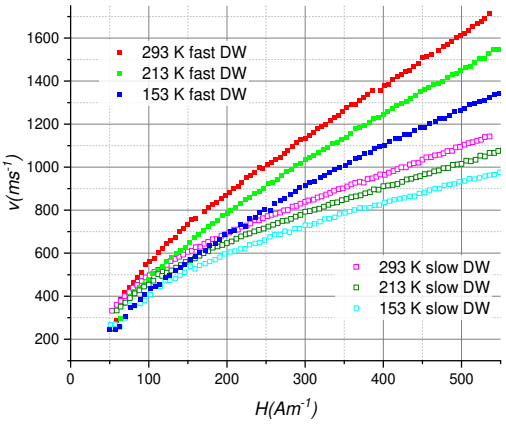


Fig. 2. Dependences of DW velocity on axial magnetic field obtained for various values of temperature applied to the sample for slow (open symbols) and fast (solid symbols) DWs.

From the measured time-dependent voltage induced in PuC due to the propagating DW, the local DW velocity as a function of wall position *x* in the PuC can be calculated [13].

When the DW propagates through a region of rapid change (region of application of increasing magnetic field) in the driving magnetic field a sharp DW velocity peak was observed [13]. We wanted to compare how the observed peak reacts when the temperature changes. In Fig. 3 we can observe a comparison of these peaks. Since the *v(H)* dependences decrease their steepness, we could not compare this rapid change when the temperature changes under the same conditions (same velocity).

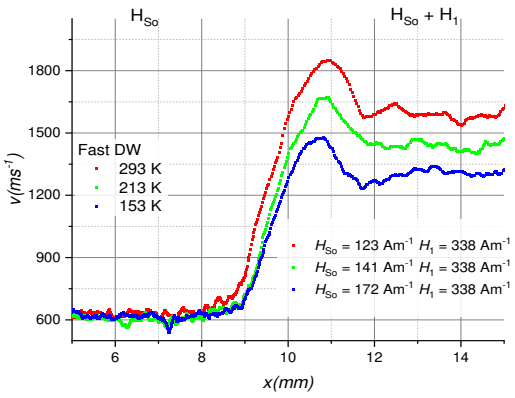


Fig. 3. Fast DW velocity vs DW position in PuC for change in DW velocity at three different temperatures.

It should be noted, that the peak does not change its parameters under the influence of temperature. We see that when there is a smaller change in velocity (smaller difference between initial and final velocity – Fig. 3), the peak is smaller.

III. CONCLUSION

This paper presents an experiment and processed data which provide information about changes in the velocity of a DW propagating along a microwire in a region of rapid change in the applied magnetic field under influence of temperature. The effects of temperature on DW velocity vs. axial magnetic field dependences were studied in a glass-coated microwire sample with a strong unidirectional effect. A sharp peak in DW velocity was observed as DW passed through a region of the increased magnetic field. We encountered a problem when comparing these peaks as we have different velocity differences at different temperatures. We need to ensure a rapid change in velocity (same initial and final velocity) during the measurement so that we can compare it.

ACKNOWLEDGMENT

This research was supported by VEGA grant No. 1/0250/21 from the Scientific Grant Agency of the Ministry for Education of the Slovak Republic. The paper has been prepared under the support of Grant FEI-2023-92.

REFERENCES

[1] Q. Wang, M. Li, X. Niu, M. Liu, B. Wang Model and Design of High-Temperature Ultrasonic Sensors for Detecting Position and Temperature Based on Iron-Based Magnetostrictive Wires IEEE Sens. J., 21 (23) (2021), pp. 26868-26877, 10.1109/JSEN.2021.3119895

[2] C. Tan, Z. Dong, Y. Li, H. Zhao, X. Huang, Z. Zou, J.-W. Jiang, Y.-Z. Long, Jiang, T.-Y. Zhang, B. Sun, A high performance wearable strain sensor with advanced thermal management for motion monitoring, Nat. Commun. 11(1) (2020) 3530, 10.1038/s41467-020-17301-6.

[3] Z. Liu, T. Zhu, J. Wang, Z. Zheng, Y. Li, J. Li, Y. Lai Functionalized Fiber-Based Strain Sensors: Pathway to Next-Generation Wearable Electronics Nano-Micro Lett., 14 (2022), p. 61, 10.1007/s40820-022-00806-8

[4] H.B. Motra, J. Hildebrand, A. Dimmig-Osburg Assessment of strain measurement techniques to characterise mechanical properties of structural steel Eng. Sci. Technol. an Int. J., 17 (2014), pp. 260-269, 10.1016/j.jestch.2014.07.006

[5] M. Albany, E. Alsahafi, I. Alruwili, S. Elkhediri A review: Secure Internet of thing System for Smart Houses Procedia Comput. Sci., 201 (2022), pp. 437-444, 10.1016/j.procs.2022.03.057

[6] H.Chiriac, T. A. Ovari, Amorphous glass covered magnetic wires: preparation, properties, applications, Prog. Mater. Sci. 40 (1996), pp. 333–407.

[7] Jiles, D. C. Recent advances and future directions in magnetic materials. Acta Mater. 51, 5907–5939 (2003).

[8] Samuhel, S., Horniaková, J., Duranka, P., Onufer, J., & Ziman, J. (2021, November). Dynamics of domain wall in rapidly-changing magnetic field. In AIP Conference Proceedings (Vol. 2411, No. 1, p. 050013). AIP Publishing LLC.

[9] J. Onufer, J. Ziman, P. Duranka, J. Kravčák, "The Study of Closure Domain Structure Dynamics in Bista-ble Microwires Using the Technique of Three-Level Field Pulses", IEEE Trans. Magn., vol. 55 (2019), 2000106.

[10] Onufer, J., Ziman, J., & Kladivová, M. (2015). Unidirectional effect in domain wall propagation observed in bistable glass-coated microwire. Journal of Magnetism and Magnetic Materials, 396, 313-317.

[11] Onufer, J., Ziman, J., Duranka, P., & Kladivová, M. (2018). The influence of annealing on domain wall propagation in bistable amorphous microwire with unidirectional effect. Physica B: Condensed Matter, 540, 58-64.

[12] Horniakova, J., Onufer, J., Ziman, J., Duranka, P., & Samuhel, S. (2021). Changes in geometry of propagating domain wall in magnetic glass-coated bistable microwire. Journal of Magnetism and Magnetic Materials, 529, 167846.

[13] J. Onufer, J. Ziman, P. Duranka, S. Samuhel, J. Horniaková and M. Kladivová, "Dynamics of Single Domain Wall Propagating in Bistable Microwire in Rapidly Changing Magnetic Field," in IEEE Transactions on Magnetics, vol. 58, no. 11, pp. 1-6, Nov. 2022, Art no. 2001506, doi: 10.1109/TMAG.2022.3207691.

Overview: Explainability of Deep Learning Models in Medical Data

[1]*Michal KOLÁRIK (3rd year),*
Supervisor: [2]Ján PARALIČ, Consultant: [3]Martin SARNOVSKÝ

[1,2,3]Dept. of Cybernetics and Artificial Intelligence, FEEI TU of Košice, Slovak Republic

[1]michal.kolarik@tuke.sk, [2]jan.paralic@tuke.sk, [3]martin.sarnovsky@tuke.sk

***Abstract*—Deep learning is increasingly finding its application in medical tasks. However, this also brings new challenges. Certainly, deep learning has the potential to revolutionize healthcare and improve patient outcomes. However, to achieve this potential, it is critical to address the challenges of privacy, interpretability, and trustworthiness of AI models in medicine. This paper presents an overview of publications that address some of the challenges in using AI in medicine. One of the problems is ensuring the privacy of patient data. Other articles address the use of explainable methods for deep learning models to explain the model and increase trust in AI models in the medical field.**

***Keywords*—Explainability, XAI, data synthetizing, ultrasonography, ECG**

I. Introduction

Deep learning methods are a current trend in data processing. However, in the medical field, it is important to be able to describe and explain the model's behavior under different situations, in addition to the actual performance of the model. Moreover, this trend is also followed by legislation that requires AI models' decisions to be safe, ethically fair, and transparent. The GDPR law offers us the right to an explanation of the model when dealing with our data. In the medical field, handling personal data is even more critical. The explainability of artificial intelligence (XAI) is a type of artificial intelligence that can provide understandable explanations about its decision-making processes. XAI is designed to make AI systems more transparent and interpretable to humans, so that we can understand how they arrived at a particular decision.

This paper presents an overview of my recently published papers in the medical domain and the current state of my dissertation work. Most of the work focuses on using deep neural networks and their explainability but also on problems such as anonymization and synthesis of medical reports [1].

The papers [2] and [3] focus on the use of explainability methods applied to deep learning models in different fields of medicine and different types of data, like ECG heart sensory data or CT lung images with COVID-19. At the same time, we also published a survey paper [4] that focuses on video processing problems in the medical domain, such as solutions using explainable methods or a survey of freely available datasets for medical videos.

This paper presents brief summary of the main research results from our journal publications indexed in Current Contents Connect database [4], [2], the peer-reviewed journal [1], and a publication at the international conference CINTI 2022 [3].

II. Review

A. Principles of Synthetizing Medical Datasets

In paper [1] we have addressed the issue of medical data privacy. Working with personal and medical data requires meeting legislative requirements and secure data handling, for example, by using anonymization or synthesis. In this work, we have made an overview of methods for data anonymization and synthesis. Based on the analysis of the options for data synthesis, we chose the GAN method, and its CTGAN [5] architecture suitable for tabular data. We explored the data synthesis options on three different sized datasets from the medical domain. For evaluation, we used quantitative metrics for the similarity of the datasets Kolmogorov-Smirnov - KSTest [6] and Chi-squared - CSTest [7] as well as visualization of the datasets in the form of correlation matrices. We evaluated the quality of the synthetic data with a decision tree model whose metrics were compared to the original and synthetic data. As a result, on the cardiovascular disease dataset, we obtained F1 score metrics for the original dataset of 0.73 and for the synthetic dataset 0.72. For the tree trained on the synthetic dataset and tested on the original dataset score was 0.70. The results show that by synthesizing the data, it is possible to approximate the original dataset very well and achieve similar quality results for the prediction of machine learning models.

B. Explainability of deep learning models in medical image classification

The possibilities offered by deep neural networks in image data analysis can also be very beneficial in the field of medicine. The requirements for medical AI applications tend to be more demanding than in other fields and require the model to be safe to use and as accurate as possible. We can try to explain the model decisions to identify problems and increase trust in AI models. There are now several explainability methods that can, for example, highlight the areas in the image that had the most influence on the model's decision. In the work [3], we focused on the possibility of using CNN and DenseNet [8] network for CT (Computer tomography) lung classification to detect patients with COVID-19. We used public dataset available at Kaggle platform the SARS-COV-2 Ct-Scan Dataset [9]. The CNN model achieved an accuracy of 0.9296 and F1 score of 0.9281, while pre-trained DenseNet

achieved results with an accuracy of 0.9960 and F1 score of 0.9955. We applied several explainability methods, namely Lime, Vannilla Gradient, and SmoothGrad, to our models in their standard form. To evaluate the explanatory power of the models, we approached domain experts and provided them with a questionnaire with sample outputs of the XAI methods. An interesting outcome of this evaluation was that the physicians focused on different domains of the CT scan in their decision-making than the model focused on. These results offer important information for future model building and show that the model can learn decisions based on different information than humans. In some cases, AI can alert the physician to important areas of the scan that the physician would not otherwise focus on.

C. Perturbation-based explainable AI for ECG sensor data

In the work [2], we have focused on exploring explainable models in the domain of stream sensor data. A lack of research focuses on XAI methods for this type of data. We focused our research on the CinC 2021 (Computing in Cardiology Challenge 2021) medical dataset. The dataset contains 10,344 labeled patient records, while it is 12-lead ECG data.

By analyzing the current state of the art, we found that ECG data is often processed by CNN architectures that take the ECG data on input as image. We chose to work directly with raw time-series data, using a 1D CNN architecture sufficient for this type of data. The simple architecture decreases the computational time of the proposed post hoc explainability method over 2D CNN architectures. However, the main contribution of this work was the design and implementation of a perturbation-based XAI method. The output of this method was the visualization of the most important regions of the ECG waveform that had the most significant influence on the model decision. We had the results of our work evaluated by domain experts (five final-year medical students). The result of the questionnaire confirmed that adding explainability to the solution increased their trust in the system and that XAI methods are important in medicine. However, several participants were concerned that using only the ECG recording was insufficient. This is because many other records are used in practice, such as patient records, test results, etc.

D. Explainability of deep learning models in medical video analysis: a survey

In the survey paper [4], we offered an overview of recent approaches for using explainable methods in deep learning on video processing tasks in the medical domain. The paper introduces the XAI domain and summarizes the most critical requirements of explainability in medical applications. Next, we present an overview of existing methods and metrics for explainability evaluation, mainly focused on the application area of video data processing in the medical domain. We provided an overview of recent works that use XAI methods in video analysis and freely available datasets used in the works. Finally, we identified open research issues in this area. We intend to focus on some of these in our future work.

III. Work in progress

We continue investigating explainable methods on lung USG image and video data. Physicians have collected and annotated the USG patient data over the past year. We currently have labeled videos for different classification and segmentation tasks, such as A-lines, B-lines, lung sliding, lung point, or effusion detection.

In work focused on A-lines detection, we use the YOLOv5 [10] architecture, where we use the EigenCAM [11] method in their standard form to represent the region important for the decision model. However, in initial experiments, we found that even when the model labels the correct object in the image, it does not always focus only on the immediate surroundings of the object (which is the expected state). By gradually fine-tuning the model, we eliminated this problem while improving the model's performance. An example of the YOLO model prediction can be seen in Fig. 1. XAI techniques can help to identify potential biases or errors in the model's decision-making that may have been introduced during training or implementation. By making the YOLO model more transparent and interpretable, XAI techniques can help to ensure that it is making fair and unbiased decisions.

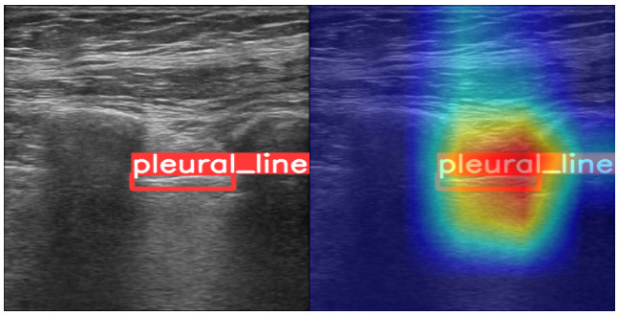


Fig. 1. Results of YOLOv5 model. Left: object of pleural line detected in image. Right: results of EigenCAM method for model explanations.

We further exploit the prediction from the output of the YOLOv5 model to obtain the region of the USG record on which the pleural line is located. If we select a 1px image column in this region and pass, e.g., 128 frames of video, we obtain a so-called B-Model image, e.g., an image of 128 pixels wide in the time domain of the video. The images thus obtained form a new dataset for us, in which the attributes are visible over time. This dataset is suitable, for example, for the classification of lung motion and absence of lung motion.

On this dataset, we are currently investigating a self-explaining xDNN [12] network, which works on the principle of assigning a predicted image to a class based on the similarity of images (or the comparison of parameters for given images in the layers of the neural network). This principle allows us to explain the model's decisions in a way that is understandable to the average person, e.g.: "I decided to assign an image to a class because it most closely resembles an image I already know, and it is this one." The user can directly judge whether a given image is similar to the predicted one or whether they are very different.

IV. Conclusions

My dissertation focuses on selected aspects of explainable artificial intelligence methods in a medical data environment. One of the aspects that needed to be explored in this context was the legislative requirements for working with personal and medical data. Another area I explored in this thesis was the use of explainable methods in distinct medical tasks. Although my

primary interest is in image and video data, I also addressed the issue of using XAI methods with time series and ECG sensor data. To evaluate the explainability of the models, I have always tried to consult the outputs of the methods with domain experts, and I have applied the knowledge gained in other publications. My work continues to explore new approaches to explain models that would address the shortcomings of using current methods. I believe the output of my work will be useful in future medical applications.

Acknowledgment

This work was partially supported by the Scientific Grant Agency of the Ministry of Education of the Slovak Republic and Academy of Science of the Slovak Republic under grant no. 1/0685/21 and partially by The Slovak Research and Development Agency under grants no. APVV-20-0232 and APVV-17-0550.

References

[1] M. Kolárik, L. Gojdičová, and J. Paralič, "Principles of synthesizing medical datasets," *Acta Electrotechnica et Informatica*, vol. 22, no. 4, pp. 25–29, 2022. [Online]. Available: https://doi.org/10.2478/aei-2022-0019

[2] J. Paralič, M. Kolárik, Z. Paraličová, O. Lohaj, and A. Jozefík, "Perturbation-based explainable ai for ecg sensor data," *Applied Sciences*, vol. 13, no. 3, 2023. [Online]. Available: https://www.mdpi.com/2076-3417/13/3/1805

[3] M. Kolarik, M. Sarnovsky, J. Paralic, and P. Butka, "Explainability of deep learning models in medical image classification," in *2022 IEEE 22nd International Symposium on Computational Intelligence and Informatics and 8th IEEE International Conference on Recent Achievements in Mechatronics, Automation, Computer Science and Robotics (CINTI-MACRo)*, 2022, pp. 000 233–000 238.

[4] M. Kolárik, M. Sarnovský, J. Paralič, and F. Babič, "Explainability of deep learning models in medical video analysis: a survey," *accepted for publication in PeerJ Computer Science*.

[5] L. Xu and K. Veeramachaneni, "Synthesizing tabular data using generative adversarial networks," 2018. [Online]. Available: https://arxiv.org/abs/1811.11264

[6] C. Kang, "Is synthetic dataset reliable for benchmarking generalizable person re-identification?" 2022. [Online]. Available: https://arxiv.org/abs/2209.05047

[7] S. Bourou, A. El Saer, T.-H. Velivassaki, A. Voulkidis, and T. Zahariadis, "A review of tabular data synthesis using gans on an ids dataset," *Information*, vol. 12, no. 9, 2021. [Online]. Available: https://www.mdpi.com/2078-2489/12/9/375

[8] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *CVPR*. IEEE Computer Society, 2017, pp. 2261–2269. [Online]. Available: http://dblp.uni-trier.de/db/conf/cvpr/cvpr2017.html

[9] E. Soares, P. Angelov, S. Biaso, M. H. Froes, and D. K. Abe, "Sars-cov-2 ct-scan dataset: A large dataset of real patients ct scans for sars-cov-2 identification," *medRxiv*, 2020. [Online]. Available: https://www.medrxiv.org/content/early/2020/05/14/2020.04.24.20078584

[10] G. Jocher, A. Chaurasia, A. Stoken, J. Borovec, NanoCode012, Y. Kwon, K. Michael, TaoXie, J. Fang, imyhxy, Lorna, Yifu), C. Wong, A. V, D. Montes, Z. Wang, C. Fati, J. Nadar, Laughing, UnglvKitDe, V. Sonck, tkianai, yxNONG, P. Skalski, A. Hogan, D. Nair, M. Strobel, and M. Jain, "ultralytics/yolov5: v7.0 - YOLOv5 SOTA Realtime Instance Segmentation," Nov. 2022. [Online]. Available: https://doi.org/10.5281/zenodo.7347926

[11] M. B. Muhammad and M. Yeasin, "Eigen-CAM: Class activation map using principal components," in *2020 International Joint Conference on Neural Networks (IJCNN)*. IEEE, jul 2020. [Online]. Available: https://doi.org/10.1109%2Fijcnn48605.2020.9206626

[12] P. Angelov and E. Soares, "Towards explainable deep neural networks (xdnn)," 2019.

Model Predictive Control in Electrical Drives: An Overview

[1]*Lukáš PANCURÁK (1st year),*
Supervisor: [2]*Karol KYSLAN*

[1,2]Dept. of Electrical Engineering and Mechatronics, FEI TU of Košice, Slovak Republic

[1]lukas.pancurak@tuke.sk, [2]karol.kyslan@tuke.sk

***Abstract*—Model predictive control (MPC) has gained widespread acceptance as an advanced control technique, not only in academia but also in industrial environment. MPC has become increasingly popular in the field of motor control over the past two decades, mostly due to its reliability, flexibility, potential to increase efficiency by reducing switching losses, etc. Recently, MPC capabilities and qualities have been improved and enhanced, and new control topologies have emerged. This paper presents an introduction into model predictive control methods suitable for applications in electrical drives, their fundamental principles, analyse the current state of MPC control, and examine innovative solutions to existing challenges.**

***Keywords*—Electric drive control, finite control set, model predictive control, motor control**

I. Introduction

One of the first applications of MPC in power electronics can already be found at the beginning of 1980s. [1] However, the control algorithm requires a large amount of calculations in relatively short time periods, which at the time meant unsatisfactory results for motor control, mostly due to lack of computational power. In recent years, popularity of MPC in power electronics and motor control has increased considerably, thanks to exponential increase in computational capabilities of microprocessors. Furthermore, higher dynamic performance, potential to include non-linearities, flexible and intuitive design consisting of a predictor based on the system model and the cost function increases interest of the scientific community as well.

Nowadays, a requirement of precise motor control is realized almost exclusively through the power converters with semiconductor power switches, mostly IGBT or MOSFET transistors. A variety of techniques are used to control the states of the power switches, enabling a converter to generate voltage with adjustable amplitude and frequency. The most commonly used control method today is Field Oriented Control (FOC) using a linear regulator, which control the switching states on a converter via PWM or SVM modulation [2]. At its core, the approach using linear PI controllers on a non-linear system. Compared to linear regulators, design of an MPC allows to include non-linearities and to control switching states directly without a need for a modulator.

This paper is an overview of MPC methods within the context of motor control. More specifically, its aim is to summarize current state of the research in predictive control of electrical drives, assess major advantages and disadvantages of MPC and to present an MPC as effective and reliable controller to deal with modern issues and challenges in motor control.

II. Predictive control methods

Predictive control has numerous applications and has gained popularity in the realm of power electronics relatively recently. The key aspect of all predictive control methods is the utilization of a model of the controlled system. This model enables the controller to predict future behaviour and states of controlled variables. This information is further used by the controller to evaluate and apply the optimal actuation inputs based on predefined optimization criteria.

In literature, the categorization of predictive control is not uniform and the boundaries between what is considered predictive control and what is not are vague. Anyway, from perspective of motor control one of the ways to categorize methods of predictive control are by ones that does require modulator and ones that do not. A way to categorize predictive control algorithms is provided in Fig. 1.

Deadbeat Predictive Control (DPC) achieves optimal actuation by using system model to calculate required reference voltage to ensure zero error in the next sampling interval. Calculated value of reference voltage is applied using a modulator [3]. Research in DPC achieved promising results. Today's DPC provides steady-state reference tracking performance and low computational demands that is on par with PI control, and far superior to Finite control set MPC (FCS-MPC) [4]. However, DPC lacks in dynamics, efficiency, and all-around performance of the system, compared to FCS-MPC.

The optimization criterion for hysteresis-based predictive control (HBPC) is maintaining the controlled variable within the area of hysteresis. This control methods doesn't require a modulator and can essentially be classified as variant of FCS-MPC. In recent years, research in HBPC shown promising result mostly in high-power generator applications as inverters output voltage regulator [5].

Similarly, in trajectory-based control, the controlled variable is forced to follow a pre-defined trajectory. This method has found its application mostly in microgrid systems, where it helps to minimize total harmonic distortion and decrease root-mean-square error between reference and controlled value[6].

In the case of MPC, the optimization criterion is defined using a cost function, which allows for higher flexibility in the design of the controller, and the possibility of controlling multiple variables with one controller. In the domain of motor

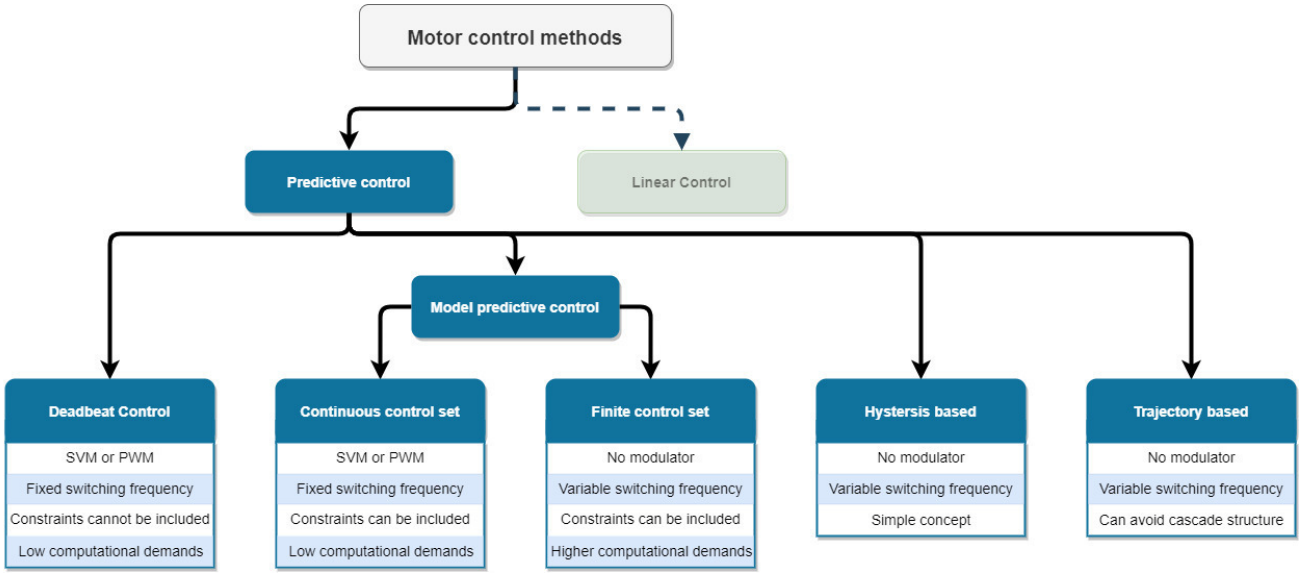


Fig. 1: General overview of predictive control methods applied in electrical drives

control, the MPC offers exceptional dynamics and efficiency, making it a highly attractive control method for demanding drives. Furthermore, excellent flexibility allowed to effectively implement MPC to variety of electric drives from six-phase induction motor [7] to three-phase PMSM with position control [8]. Although MPC controllers can be implemented in complex systems, issues may still arise, such as high-frequency ripples in the current, among others.

III. Model predictive control

A common feature of all types of MPC controllers is that they use system model to predict future behaviour and states of the system. The prediction's time horizon is predefined by design, but often restricted by the desired dynamics of the regulated system and also by the sampling frequency of the microprocessor.

MPC design approaches can be divided into two groups: Continuous Control Set (CCS) MPC and Finite Control Set (FCS) MPC. Both approaches has the ability to predict future behaviour of the system, and then selects the control input that minimizes a predefined cost function, which leads to optimal control actuation that will drive the system toward desired goal. In CCS-MPC the control input is typically a continuous variable, which can take any value within specified range. The controller uses optimization algorithm to find the optimal control input that minimizes a predefined cost function. Optimal actuation is then applied using a modulator. In contrast, FCS-MPC has finite number of discrete control inputs. In electrical drive applications, number of control inputs corresponds to the number of inverter switching sequences. This method does not require a modulator, as the actuations are directly applied to the converter in the form of a switching sequence.

A. Finite Control Set Model Predictive Control

Most significant advantage of FCS-MPC, and also one of the reason why it is still relevant and interesting topic of today, is the potential to increase the efficiency and reduce the energy consumptions of electric drives, particularly for high-power motor applications [9]. Control techniques like FCS-MPC and The FCS-MPC reduce the inverter's switching frequency, leading to lower losses and improved inverter efficiency [10]. Another significant advantages of MPC are:

- Dynamic performance: Improved dynamic performance compared to conventional control techniques, resulting in faster and smoother response to changes in the system's operating conditions.
- Versatility: FCS-MPC can be applied to wide range of applications.
- Multivariable design: Possibility to include multiple variables within a single cost function.
- Robustness: Performance remains stable even when operating conditions change. Furthermore, the use of a mathematical model of the system facilitates the controller's ability to adapt to varying conditions. This results in high robustness, stability and reliability.
- Nonlinearity handling: It is very simple to include nonlinearities into the cost function.
- Expandability: The MPC controller is suitable for further expansion and modification if necessary.

The drawbacks of MPC controllers comes in form of high demand on computation power. The quality of control is directly linked to the quality of system's model, which requires complex calculations in the microprocessor. When operating conditions and system's parameters are time-varying, an adapting algorithm might be necessary. This further increases computing requirements. [11] Another disadvantages of MPC are:

- When controlling a large number of variables, it can become challenging to monitor each of them separately, due to their interrelated nature. One of the solutions is to use multiple MPC controllers, though this increases computing demands [12].
- Due to variable switching frequency of FCS-MPC, spectrum of frequencies where higher harmonic distortions can occur is significantly wider compared to using PWM or SVM. This results in complicated signal filtering [13]-
- Sensitivity to disturbances: MPC can be sensitive to disturbances, measurement noise, and model uncertainties. Simple solution to this issue is to incorporate disturbances and measurement noise into the model or to mitigate

their impact by using a filter. Due to the noisy nature of speed measurement it is necessary to implement observer. Using the Kalman Filter as an observer is beneficial due to its attributes that enable it to meet both requirements effectively.

The FCS-MPC control the power switches and switching sequence directly. Depending on the optimization method, FCS-MPC can be further divided into Optimal switching vector (OSV-MPC) [14] and Optimal switching sequence (OSS-MPC) [15] algorithms. When designing an MPC controller it is important to consider the type of power converter. For a three-phase motor, the most commonly used converter is the three-phase, two-level converter with six power switches, commonly known as voltage source inverter (VSI). It has eight switching states in total. For OSV-MPC each state generates their own voltage vector, which includes six active voltage vectors (U_1 - U_6), each composed of a unique combination of switched power switches (S1 - S6), and two zero voltage vectors (U_0 and U_7), as shown in Fig. 2. The main drawback of this method is that it only uses a single output voltage vector throughout the entire switching cycle, which can lead to repeated use of the same output voltage vector over multiple consecutive switching cycles. This generates variable switching frequency with wide range of harmonic disturbances [16].

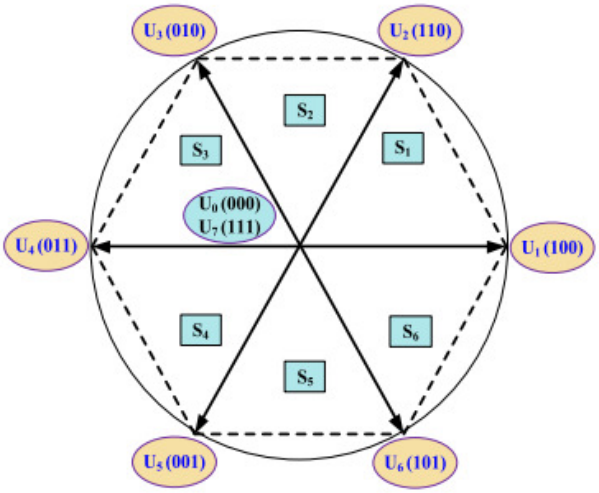


Fig. 2: Voltage Vectors of VSI

The OSS-MPC approach overcomes this issue by implementing control set composed of limited number of possible switching sequences per switching period. In this way, OSS-MPC takes the time into account as an additional decision variable, which in a way resembles a modulator.

B. System Model

A model of a controlled system is necessary for implementation of MPC controller. It is used to predict future behaviour based on current inputs and measurements. The quality of the predictions depends on the accuracy of the system model, and therefore it is critical to have a good understanding of the controlled system. The MPC model can be linear or non-linear, and it can be derived from first principles or obtained by system identification methods.

If the model of a controlled system is available, it has to be discretized for its further implementation into microcontroller. Discretization techniques can affect quality and complexity of the model. Typically, this is achieved through the use of an Euler approximation or a Taylor series expansion. Recently, study [17] has shown new discretization method based on the input-state linearization of non-linear control. Other studies also explored the use of model-free predictive control [18]. The results indicated that the prior knowledge of the physical model of the inverter and the load is not a necessity for the implementation of predictive control. In addition, these results confirm that this strategy is much less susceptible to model inaccuracies compared to regular MPC.

C. Cost Function

As was stated before, the optimal voltage vector is selected by minimizing the cost function. Cost function design and complexity is dependent on number of controlled variables and constraints. The majority of terms in a cost function represent the deviation of a variable from a reference. In other words, it is the absolute, square, or integral value of an error of controlled variable. To limit or reduce control effort an actuation constraints can be implemented. In drive applications, the control effort is associated with voltage, current or switching frequency variations. Thus, it is crucial to reach a compromise between control effort and optimal operation [19]. The fundamental cost function with constraints can be expressed as

$$g = \lambda_1 |x_{ref}(k) - x_p(k)|^2 + \lambda_2 |y(k-1) - y(k)|, \quad (1)$$

where x_{ref} is the reference value of controlled variable, x_p is predicted value of controlled variable, λ is a weighting factor that allows the level of compromise to be adjusted. The second part of cost function is actuation constraint where y is actuation. This part is usually expressed in form of a more complex function. There is no typical way of determining the value of weighting factors. The most basic approach is to assign random values and tune them depending on simulation results, what is known as trial-and-error method. Usually, higher values are assigned to weighting factors of variables with higher time constant.

A significant attention in research is focused on optimizing the cost function and determining the optimal weighting factors. Weighting factor selection techniques can be split into offline and online methods, and further divided into multiple groups, as shown in Fig. 3.

Recently, there have been promising attempts at offline optimization of a cost function utilizing artificial neural network [20]. There were also attempts at online optimization

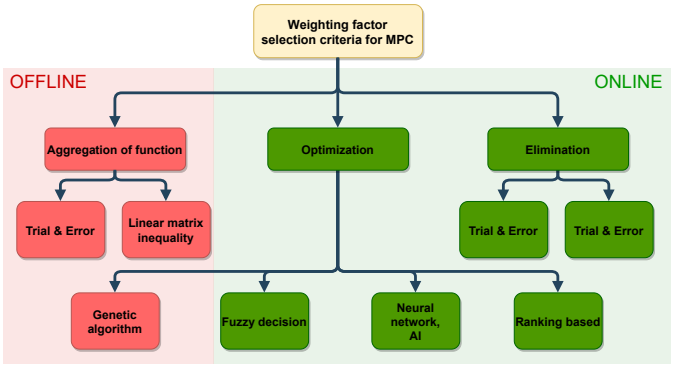


Fig. 3: Weighting factor selection methods used in MPC.

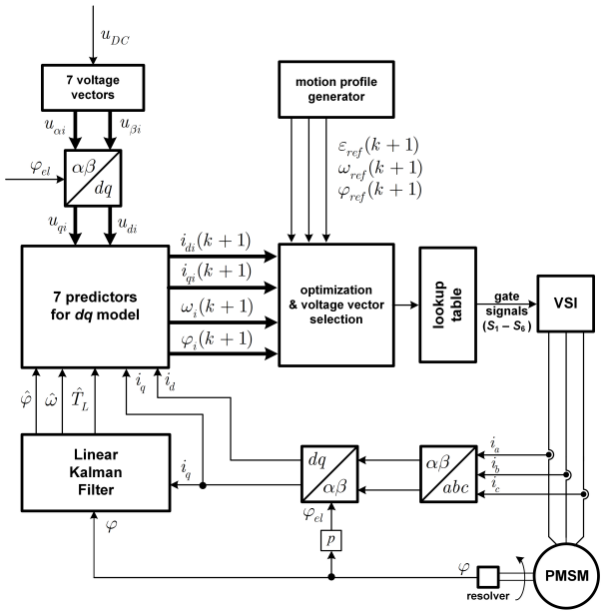


Fig. 4: Control Scheme of Feedforward FCS-MPC [8].

by improving MPC design to reduce torque ripples [21]. Further improvements was achieved with online weight tuning using fuzzy [22], which however can be impractical due to already high computational demands. Another more promising online optimization strategy replaces the minimization of a cost function with a multi-objective optimization using fuzzy decision making [23]. The results demonstrate that this strategy compares favourably with conventional predictive control approaches in terms of harmonic disturbances, torque ripples, and computational demands.

D. Control Structure

Final control scheme consist of predictor with a discretized system model, optimization part where cost function is minimized and voltage vector is selected. The example of such control scheme proposed in [8] is shown in Fig. 4. It was used for position control of permanent magnet synchronous machine (PMSM). It used linear Kalman filter (LKF) for estimation of unknown load torque and for estimation of drive speed with high sampling rate. A generation of reference trajectories for a drive was provided by standard motion profile generator. Weighting factors were selected by trial-and-error method. It was shown that proposed FCS-MPC controller is able to track reference position with satisfactory accuracy.

IV. Future efforts

The crucial components of any MPC strategy applied to power converters and drives are the prediction model, cost function, and optimization algorithm, as determined through analysis. Despite extensive research efforts in the fields predictive control strategies applied to power converters and drives, various challenges and limitations have been recognized. Although progress has been made in addressing some of these issues, others remain unresolved and require ongoing research efforts. Currently, our focus is to improve existing algorithms of FCS-MPC, by implementing optimization algorithm that would not significantly increase computational demands. Moreover, we will try to utilize the flexibility and expendability of the MPC to efficiently control electric drives powered by battery system.

Acknowledgment

This work was supported by Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovakia under Grant FEI-2023-98 *Experimentálne overenie prediktívnej regulácie synchrónneho motora s permanentnými magnetmi*.

References

[1] J. Holtz and S. Stadtfeld, "A predictive controller for the stator current vector of ac machines fed from switched voltage source," in *International Power Electronics Conference, IPEC, Tokyo, pp. 1665–1675*, 1983.

[2] P. Cortes, M. P. Kazmierkowski, R. M. Kennel, D. E. Quevedo, and J. Rodriguez, "Predictive control in power electronics and drives," *IEEE Transactions on Industrial Electronics*, vol. 55, no. 12, pp. 4312–4324, 2008.

[3] T. Kawabata, T. Miyashita, and Y. Yamamoto, "Dead beat control of three phase pwm inverter," *IEEE Transactions on Power Electronics*, vol. 5, no. 1, pp. 21–28, 1990.

[4] X. Zhang, B. Hou, and Y. Mei, "Deadbeat predictive current control of permanent-magnet synchronous motors with stator current and disturbance observer," *IEEE Transactions on Power Electronics*, vol. 32, no. 5, pp. 3818–3834, 2017.

[5] X. Zhang, Y. Wang, C. Yu, L. Guo, and R. Cao, "Hysteresis model predictive control for high-power grid-connected inverters with output lcl filter," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 1, pp. 246–256, 2016.

[6] J. Yu, F. Gao, and J. Xu, "Trajectory based model predictive control strategy for lc filtered vsis," in *2022 IEEE PES Innovative Smart Grid Technologies - Asia (ISGT Asia)*, 2022, pp. 570–574.

[7] I. Gonzalez-Prieto, M. J. Duran, J. J. Aciego, C. Martin, and F. Barrero, "Model predictive control of six-phase induction motor drives using virtual voltage vectors," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 1, pp. 27–37, 2018.

[8] K. Kyslan, V. Šlapák, F. Ďurovský, V. Fedák, and S. Padmanaban, "Feedforward finite control set model predictive position control of pmsm," in *2018 IEEE 18th International Power Electronics and Motion Control Conference (PEMC)*, 2018, pp. 549–555.

[9] T. Li, X. Sun, G. Lei, Y. Guo, Z. Yang, and J. Zhu, "Finite-control-set model predictive control of permanent magnet synchronous motor drive systems—an overview," *IEEE/CAA Journal of Automatica Sinica*, vol. 9, no. 12, pp. 2087–2105, 2022.

[10] R. Baidya, R. P. Aguilera, P. Acuña, T. Geyer, R. A. Delgado, D. E. Quevedo, and H. d. T. Mouton, "Enabling multistep model predictive control for transient operation of power converters," *IEEE Open Journal of the Industrial Electronics Society*, vol. 1, pp. 284–297, 2020.

[11] J. Rodriguez and P. Cortes, *Predictive Control of Power Converters and Electrical Drives*. John Willey & Sons, 2012.

[12] M. G. Forbes, R. S. Patwardhan, H. Hamadah, and R. B. Gopaluni, "Model predictive control in industry: Challenges and opportunities," *IFAC-PapersOnLine*, vol. 48, no. 8, pp. 531–538, 2015, 9th IFAC Symposium on Advanced Control of Chemical Processes ADCHEM 2015.

[13] S. Borreggine, V. G. Monopoli, G. Rizzello, D. Naso, F. Cupertino, and R. Consoletti, "A review on model predictive control and its applications in power electronics," in *2019 AEIT International Conference of Electrical and Electronic Technologies for Automotive (AEIT AUTOMOTIVE)*, 2019, pp. 1–6.

[14] T. Geyer and D. E. Quevedo, "Multistep finite control set model predictive control for power electronics," *IEEE Transactions on Power Electronics*, vol. 29, no. 12, pp. 6836–6846, 2014.

[15] S. Vazquez, A. Marquez, R. Aguilera, D. Quevedo, J. I. Leon, and L. G. Franquelo, "Predictive optimal switching sequence direct power control for grid-connected power converters," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 4, pp. 2010–2020, 2015.

[16] S. Vazquez, J. Rodriguez, M. Rivera, L. G. Franquelo, and M. Norambuena, "Model predictive control for power converters and drives: Advances and trends," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 2, pp. 935–947, 2017.

[17] F. Villarroel, J. Espinoza, M. Pérez, D. Sbárbaro, R. Ramírez, and C. Baier, "A new discretization method of model equations for predictive power converter control applications based on input-state linearization," in *IECON 2022 – 48th Annual Conference of the IEEE Industrial Electronics Society*, 2022, pp. 1–5.

[18] J. Rodríguez, R. Heydari, Z. Rafiee, H. A. Young, F. Flores-Bahamonde, and M. Shahparasti, "Model-free predictive current control of a voltage source inverter," *IEEE Access*, vol. 8, pp. 211 104–211 114, 2020.

[19] J. Rodriguez and P. Cortes, *Cost Function Selection*, 2012, pp. 145–161.

[20] C. Yao, Z. Sun, S. Xu, H. Zhang, G. Ren, and G. Ma, "Ann optimization of weighting factors using genetic algorithm for model predictive control of pmsm drives," *IEEE Transactions on Industry Applications*, vol. 58, no. 6, pp. 7346–7362, 2022.

[21] S. A. Davari, D. A. Khaburi, P. Stolze, and R. Kennel, "An improved finite control set-model predictive control (fcs-mpc) algorithm with imposed optimized weighting factor," in *Proceedings of the 2011 14th European Conference on Power Electronics and Applications*, 2011, pp. 1–10.

[22] H. Mahmoudi, M. j. Lesani, and D. Arab khabouri, "Online fuzzy tuning of weighting factor in model predictive control of pmsm," in *2013 13th Iranian Conference on Fuzzy Systems (IFSC)*, 2013, pp. 1–5.

[23] C. A. Rojas, J. R. Rodriguez, S. Kouro, and F. Villarroel, "Multiobjective fuzzy-decision-making predictive torque control for an induction motor drive," *IEEE Transactions on Power Electronics*, vol. 32, no. 8, pp. 6245–6260, 2017.

Overview of the Battery Systems and Battery Management in Electric Vehicles and a Proposal of New Modular Battery System

[1]*Daniel MARCIN (1st year)*
Supervisor: [2]Milan LACKO

[1,2]Department of Electrical Engineering and Mechatronics, FEI TU of Košice, Slovak Republic

[1]daniel.marcin@tuke.sk, [2]milan.lacko@tuke.sk

***Abstract*—The increasing demand for efficient and cost-effective battery systems has spurred research in the field of electric vehicles. The use of li-ion batteries is described, along with the battery management system and the difference between passive and active cell balancing. The future work for this system involves simulations and building a battery pack. The goal is to move the research of electric vehicles battery systems forward.**

***Keywords*—Battery, battery management system, cell balancing, electric vehicle, li-ion, modular battery system.**

I. Introduction

The popularity of electric vehicles (EVs) has risen in recent years, as they offer a solution to reduce dependence on fossil fuels and the environmental impact of transportation. The effectiveness of EVs, is highly dependent on the efficiency and reliability of its battery system (BS). The BS in an EV is a crucial factor in determining the vehicle's range, performance, and reliability. To meet the increasing demand for EVs, there is a growing need for advanced energy storage solutions that are efficient, scalable, and cost-effective [1], [2].

A modular battery system, consisting of multiple interchangeable battery modules, holds great promise for meeting the energy storage needs of electric vehicles. This system provides the flexibility to adapt to changing requirements, whether it be adjusting for user preferences or driving conditions that require varying ranges on a single charge. Another type of module can consist of supercapacitors that can be used to relieve strain on the battery while performing energy recovery braking. Additionally, the ability to add or remove modules enhances maintenance of the BS. These modular systems have the potential to fulfill the performance and cost requirements of EVs while providing the flexibility to adapt to changing needs [3], [4].

The battery management system (BMS) is an integral part of the EV BS and ensures safe operation and optimal performance. The BMS monitors the state of the cells and performs functions such as voltage and temperature monitoring, cell balancing, fault detection or protection. Cell balancing is a critical aspect of the BMS, it ensures equal distribution of charge and discharge among the cells in a battery pack. This helps to prevent overloading of individual cells, which can lead to reduced performance, increased wear, and potential safety hazards. Currently, passive cell balancing is a widely utilized technique in electric vehicles [1]. An alternative solution, active cell balancing, is more efficient and involves the transfer of energy between cells, thus avoiding energy wastage [2], [3].

This paper provides an overview of the technology used in EVs, including lithium-ion (li-ion) batteries and their required battery management systems. Passive and active cell balancing techniques and their corresponding circuit topologies are discussed, along with specific monitoring and balancing circuits. At the end, the concept of a modular battery system and its future challenges are also presented.

II. Current EVs state and battery system

The latest advancements in battery technology and manufacturing processes have made EVs more practical and accessible for a wider range of consumers. The demand for EVs is increasing rapidly, and with it, the need for longer range and more efficient battery systems. One of the key areas of research in the field of EVs is increasing the range of the battery systems. This involves developing new technologies that can store more energy in a smaller and lighter package.

Another trend in the EV battery industry is the increasing demand for efficient battery management systems. Passive balancing, which relies on the difference in self-discharge between cells, has been the standard for many years. However, as battery systems become larger and more complex, active cell balancing is becoming a more popular choice.

The demand for EVs continues to grow, however the cost of electrical energy is also increasing, which leads to a greater focus on energy efficiency and reducing the EV's overall cost.

A. Lithium-ion batteries

Li-ion batteries have become a key component in the growth of EVs. Li-ion batteries are characterized by high energy density, long cycle life, a negligible memory effect, and a low self-discharge rate. These characteristics position li-ion batteries as a highly competitive player in the global battery market, making them superior to other battery types in various applications [1].

Li-ion batteries are manufactured for EVs in cylindrical cell, prismatic or pouch cell. Cylindrical cells are the most common used type (Fig. 1). Prismatic cells are flat and rectangular, while

pouch cells (Fig. 1) are flexible and can be molded to fit different shapes. Cylindrical cells are commonly used for EVs due to their strong casing and shape that helps prevent stress on edges when assembled into a battery pack [2], [3], [4].

Fig. 1 Battery cells, battery module, battery pack [5]

The arrangement of battery cells in a battery pack is customized to meet the specific voltage and capacity needs of an electric vehicle. The cells are connected in series to reach the desired voltage, and when a higher capacity is required, additional cells are connected in parallel. This configuration which is shown in Fig. 1 allows for a battery module to be formed from multiple connected cells, and multiple battery modules to form the final battery pack [1].

B. Battery management system

Despite of li-ion batteries having a wide range of applications and numerous advantages, they can be quite unforgiving if not operated within their safe operating area (SOA) [6]. Especially, li-ion batteries used in EVs require proper safety measures to be in place, particularly due to the increased risk posed by varying driving conditions or environmental temperatures. SOA is determined by the operation range of voltage, current and temperature [7]. For li-ion cells safe operating voltage is within range from 2,4 V to 4,2 V [8]. The temperature is safe within -20 °C to 60 °C. The charge and discharge current are rated in the C- rate, which is a measure of the current relative to the cell's capacity.[9].

BMS in an EV is a crucial component responsible for ensuring the safe and efficient operation of the battery pack. In EVs, BSM primarily performs monitoring and control functions for the battery pack. BMS key functions are following [1]:

- Battery cell monitoring
- Input/output current and voltage monitoring
- Charge and discharge control
- State estimation (SOC, SOH)
- Battery protection
- Cell balancing and equalization
- Power management control
- Operating temperature control and heat management
- Communication and networking
- Data storage and acquisition
- Fault diagnosis and assessment

These functions are essential for ensuring the reliability of batteries, increasing the discharge rate, prolonging the range of EVs, maximizing power, and extending the overall lifetime of the battery pack [10], [11].

Accurate estimation of the state of the battery is a crucial aspect of BMS operation. This concerns the state of charge (SOC) and state of health (SOH). The SOC of a battery indicates the amount of charge remaining in the battery. The SOH of a battery refers to the battery's capacity compared to its original capacity when it was new. It can be represented by the amount of capacity loss or increase in resistance over time [12].

C. Cell balancing

Because of the cell-to-cell variations or different charging history, cells may be imbalanced in four ways. Cells in a series connected pack can differ from each other in terms of SOC, leakage current (self-discharge), resistance, and capacity [7]. When battery module or battery pack is imbalanced, the SOA is then limited by the cell with the lowest voltage during discharge and the highest voltage during charge, thus reducing the overall capacity of the battery pack [7].

To perform accurate balancing in a battery pack, the voltage of each cell in the series must be measured [7]. By measuring the voltage of each cell, it is possible to determine which cells are imbalanced and require balancing. There are two main types of balancing techniques used in battery management systems: passive and active cell balancing. In passive balancing method the energy is wasted in heat. Active balancing method transfers the energy from cells with higher SOC to cells with lower SOC. This method can be done by using capacitors, inductors, or transformers [13]. Balancing can be done within individual battery modules or within the entire battery pack.

Passive cell balancing utilizes dissipative resistors to remove energy from specific cells (Fig. 2). It is the simplest and cheapest balancing method. Cells with higher SOC will be selected to be discharged via shunt resistor by controlling specific switch [16]. Another way of passive balancing is by fixed resistor, without additional switch. There is continuous current passing through the resistors. Balancing of cells can be achieved without the need for active control, as the dissipative currents are proportional to the cell terminal voltages. However, the energy waste and slow pace of balancing may result in certain cells being overcharged [14].

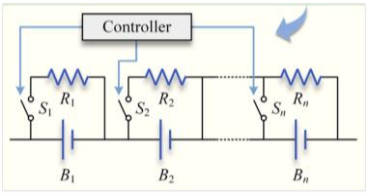


Fig. 2 Passive cell balancing using shunt resistors and switches [15]

Passive cell balancing can be operated continuously, with the resistors turning on and off as required to balance the cells in the battery pack. The effectiveness of the method depends on several factors, including the size and design of the shunt resistors, as well as the control algorithms used to manage the balancing process. Properly designed passive balancing systems can effectively maintain balance in a battery pack but may not be as precise or efficient as active cell balancing methods. One of the main disadvantages of passive cell balancing is high energy losses, which can reduce the energy efficiency of the battery pack. Despite this drawback, passive cell balancing is a widely used method in electric vehicle (EV) applications due to its simplicity, low cost, and ease of implementation [16].

Active cell balancing is another method of balancing batteries that is generally more efficient. Active cell balancing uses active electronic components, to transfer energy between cells. This allows for more precise control of the balancing process and reduces energy loss, resulting in higher energy efficiency. However, active cell balancing is typically more complex and expensive compared to passive cell balancing and may not be the best option for all applications.

- *Capacitor based active cell balancing*

From all energy storage options in active cell balancing capacitor is preferred the most. It has no magnetic components, it is small in size and cost effective [17]. The switched capacitor method is one of the most common techniques used in capacitor-based active cell balancing. There are several variations of structures used in these circuits. Single tiered switched capacitor requires 2n switches and n-1 capacitors (Fig. 3.a). This circuit has only one disadvantage which is relatively long balancing time, as only adjacent cells can be balanced directly. By adding additional tier of capacitors balancing speed can be improved because of more charge transfer paths [14], [17]. Fig. 3.b shows the structure of double tiered switched capacitor. The second tier of capacitors helps to transfer energy between nonadjacent cells in single switching cycle [17].

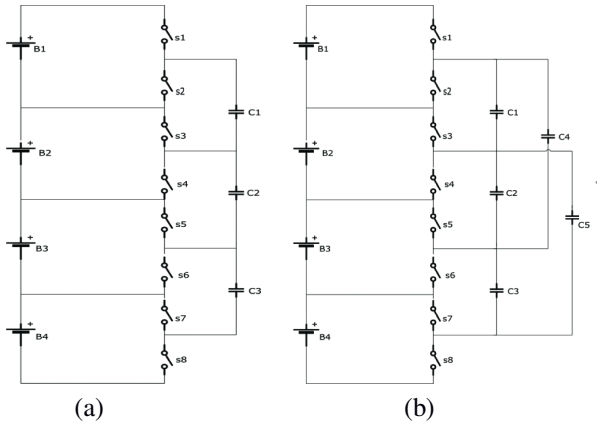


Fig. 3 Capacitor based active balancing: (a) Single tiered switched capacitor (b) Double tiered switched capacitor [14]

Different structures can be used in capacitor based active cell balancing to improve the balancing time and precision. Chain structure may be utilized to balance cells at two ends directly with additional capacitor. Parallel structure allows balancing two cells at any positions but increases the balancing circuit equivalent series resistance. Delta or mesh structures are more complex and require more components but can provide more precise and efficient balancing [14].

- *Inductor based active cell balancing*

Inductor based cell balancing is another common method used in battery balancing. Like capacitor based balancing, it also uses switches and similar circuit topologies with single or multi inductors (Fig. 4). This method is known for its fast balancing time compared to capacitor based balancing [18], [19]. However, inductors are also more complex to implement and are more expensive than capacitors.

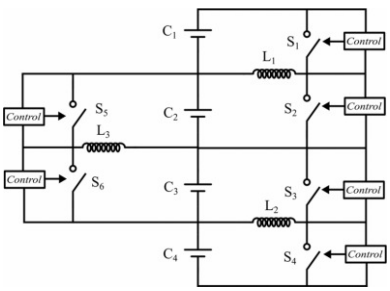


Fig. 4 Inductor based active cell balancing [19]

Basic inductor active cell balancing topology, as shown in Fig. 4, typically only allows balancing between adjacent cells [18], [19]. To balance cells that are not adjacent, more complex topologies are required, which may need additional components and increase the complexity of the system. Such as system is inductor based balancing circuit combined with capacitors or chain structure with additional capacitors [14], [20], [21].

- *Transformer based active cell balancing*

Another method of energy transfer in active cell balancing is the use of transformers. Magnetic field created by transformer is used to transfer energy from cells with higher SOC to lower SOC. A simple transformer with primary and secondary windings can be used in form of single flyback converter [14], [22]. Other way of transferring energy is by use of multi-winding transformer [23]. Primary winding of transformer is usually connected to battery pack side. Secondary winding is connected to every cell in balancing circuit [24]. Connection of multi-winding transformer is shown in Fig. 5.

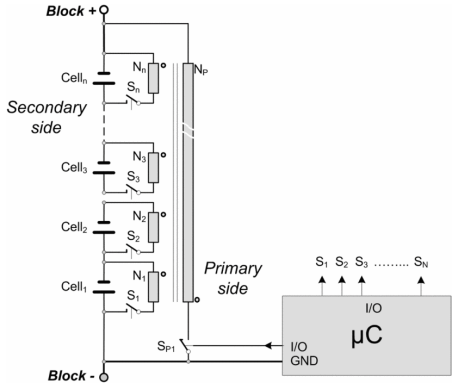


Fig. 5 Active cell balancing using transformer as flyback converter [24]

- *Other methods of active cell balancing*

Another method of active cell balancing in a battery system is to use a buck-boost converter connected to each cell. This allows for individual control of each cell and can be more effective in balancing their voltage levels. However, this method requires a separate buck-boost converter for each cell, leading to a larger scale of the balancing circuit [14]. The added complexity and size of this method may also affect its efficiency and reliability compared to other methods.

- *Monitoring and balancing circuits*

On the market there are several circuits that can be used for either monitoring or balancing cells from various manufacturers. For future purposes a monitoring circuit can be used BQ76952 from Texas Instruments. It offers monitoring from 3 to 16 cells in series [26]. Another circuit MM3513 from Minebea Mitsumi can be used with MM3474 for monitoring and balancing with capacitor active cell balancing [27]. Analog devices offer large scale of products with active cell balancing circuits like LT8584, which is for transformer based balancing as flyback converter [28].

III. Modular Battery System Proposal

Recently, there has been an increasing demand for energy-saving and cost-effective battery systems. This leads researchers to the development of new technologies in car battery systems. This also includes the development of modular

battery system which consist of interchangeable battery modules. This new technology has the potential to greatly impact the range, modularity, maintenance, or recycling of EV batteries.

Webasto recently launched their innovative "plug-and-play" battery system, offering 35 kWh of energy per battery pack. This system aims to provide a cost-efficient solution for manufacturers by eliminating the need for a separate battery system for each vehicle. The battery packs can be connected in parallel to achieve 400V with five modules or in series to achieve 800V with ten modules [25].

One of the main benefits of modular battery systems is that they allow for easier replacement of individual battery modules, rather than having to replace the entire battery pack.

Another advantage of modular battery systems is that they can be easily upgraded as new, more efficient battery technologies become available. This allows EV owners to extend the lifespan of their vehicles by simply replacing individual modules, rather than having to purchase a completely new battery pack.

These systems consist of interchangeable battery modules, which can be swapped out as needed to adjust the vehicle's battery capacity. This is particularly useful for larger vehicles like buses, as they can carry only the amount of battery needed for the specific road, reducing weight and increasing range.

IV. Future Work

The aim of this project is to design and develop a modular battery system using active cell balancing technology. The study will begin with simulations comparing active and passive cell balancing and then proceed to build three modular battery packs using cylindrical li-ion cells – approximately 10 in series. BMS for these packs will be designed using existing monitoring and balancing circuits. The design will consider communication between the battery pack and the overall vehicle's system, as well as ensuring the safety and protection of the entire battery system.

Acknowledgment

This work was supported by the Slovak Research and Development Agency under project APVV-18-0436.

References

[1] M. A. Hannan, M. M. Hoque, A. Hussain, Y. Yusof and P. J. Ker, "State-of-the-Art and Energy Management System of Lithium-Ion Batteries in Electric Vehicle Applications: Issues and Recommendations," in IEEE Access, vol. 6, pp. 19362-19378, 2018, doi: 10.1109/ACCESS.2018.2817655.

[2] Halimah, Putri Nur, Samuel Rahardian, and Bentang Arief Budiman. "Battery cells for electric vehicles." International Journal of Sustainable Transportation Technology 2.2 (2019): 54-57.

[3] Sun, Peiyi & Bisschop, Roeland & Niu, Huichang & Huang, Xinyan. (2020). A Review of Battery Fires in Electri c Vehicles. Fire Technology. 1-50. 10.1007/s10694-019-00944-3.

[4] What Types of Batteries Are Used in Electric Cars? [Online]. Available: https://dragonflyenergy.com/electric-car-batteries/

[5] Supply Chain Opportunities from Electrification of Vehicles. [Online]. Available: https://northernautoalliance.com/wp-content/uploads/2019/05/Ready-for-ACES-Supply-Chain-Opportunities-from-Electrification-of-Vehicles.pdf

[6] What is a Battery Management System? [Online]. Available: https://www.synopsys.com/glossary/what-is-a-battery-management-system.html

[7] Andrea, Davide. Battery management systems for large lithium-ion battery packs. Artech house, 2010.

[8] N. Ghaeminezhad, Q. Ouyang, X. Hu, G. Xu and Z. Wang, "Active Cell Equalization Topologies Analysis for Battery Packs: A Systematic Review," in IEEE Transactions on Power Electronics, vol. 36, no. 8, pp. 9119-9135, Aug. 2021, doi: 10.1109/TPEL.2021.3052163.

[9] Battery-Management-Systems. [Online]. Available: https://www.vde.com/resource/blob/2105544/731ed20695f9fbf2f66423407554d1c5/fact-sheet-battery-management-system-pdf-file-data.pdf

[10] Yang, Shichun, et al. Advanced Battery Management System for Electric Vehicles. Springer Nature, 2022.

[11] S. B and P. Pradeepa, "Review on Battery Management System in EV," 2022 International Conference on Intelligent Controller and Computing for Smart Power (ICICCSP), Hyderabad, India, 2022, pp. 1-4, doi: 10.1109/ICICCSP53532.2022.9862367.

[12] P. Shen, M. Ouyang, L. Lu, J. Li and X. Feng, "The Co-estimation of State of Charge, State of Health, and State of Function for Lithium-Ion Batteries in Electric Vehicles," in IEEE Transactions on Vehicular Technology, vol. 67, no. 1, pp. 92-103, Jan. 2018, doi: 10.1109/TVT.2017.2751613.

[13] L. Valda and K. Kosturik, "Comparison of Li-ion active cell balancing methods replacing passive cell balancer," 2015 International Conference on Applied Electronics (AE), Pilsen, Czech Republic, 2015, pp. 267-270.

[14] B. Jiang, Y. Liu, X. Huang and R. R. R. Prakash, "A New Battery Active Balancing Method with Supercapacitor Considering Regeneration Process," IECON 2020 The 46th Annual Conference of the IEEE Industrial Electronics Society, Singapore, 2020, pp. 2364-2369, doi: 10.1109/IECON43393.2020.9254839.

[15] Liu, Wei & Placke, Tobias & Chau, K.T.. (2021). Overview of Batteries and Battery Management for Electric Vehicles.

[16] X. Wei and B. Zhu, "The research of vehicle power Li-ion battery pack balancing method," 2009 9th International Conference on Electronic Measurement & Instruments, Beijing, China, 2009, pp. 2-498-2-502, doi: 10.1109/ICEMI.2009.5274520.

[17] R. Paul and S. S, "Electric Vehicle Cell Balancing using Single and Multi tiered Switched Capacitor," 2022 4th International Conference on Energy, Power and Environment (ICEPE), Shillong, India, 2022, pp. 1-6, doi: 10.1109/ICEPE55035.2022.9798231.

[18] A. F. Moghaddam and A. Van Den Bossche, "An Active Cell Equalization Technique for Lithium Ion Batteries Based on Inductor Balancing," 2018 9th International Conference on Mechanical and Aerospace Engineering (ICMAE), Budapest, Hungary, 2018, pp. 274-278, doi: 10.1109/ICMAE.2018.8467685.

[19] G. N, G. Yadav and G. CK, "Analysis and Implementation of Inductor Based Active Battery Cell Balancing Topology," 2020 IEEE International Conference on Power Electronics, Drives and Energy Systems (PEDES), Jaipur, India, 2020, pp. 1-6, doi: 10.1109/PEDES49360.2020.9379358.

[20] A. Panchal, K. Bhatt, S. Gitaye, M. Bhand and A. Sheikh, "Design and Simulation of an Inductor based Active Cell Balancing Circuit for Lithium-ion Batteries," 2022 4th Global Power, Energy and Communication Conference (GPECOM), Nevsehir, Turkey, 2022, pp. 89-94, doi: 10.1109/GPECOM55404.2022.9815650.

[21] T. H. Phung, J. -C. Crebier, A. Chureau, A. Collet and V. Nguyen, "Optimized structure for next-to-next balancing of series-connected lithium-ion cells," 2011 Twenty-Sixth Annual IEEE Applied Power Electronics Conference and Exposition (APEC), Fort Worth, TX, USA, 2011, pp. 1374-1381, doi: 10.1109/APEC.2011.5744771.

[22] Cui, Xiudong & Shen, Weixiang & Zhang, Yunlei & Hu, Cungang. (2017). A Novel Active Online State of Charge Based Balancing Approach for Lithium-Ion Battery Packs during Fast Charging Process in Electric Vehicles. Energies. 10. 1766. 10.3390/en10111766.

[23] Y. Shang, B. Xia, C. Zhang, N. Cui, J. Yang and C. Mi, "A Modularization Method for Battery Equalizers Using Multiwinding Transformers," in IEEE Transactions on Vehicular Technology, vol. 66, no. 10, pp. 8710-8722, Oct. 2017, doi: 10.1109/TVT.2017.2702065.

[24] C. Bonfiglio and W. Roessler, "A cost optimized battery management system with active cell balancing for lithium ion battery stacks," 2009 IEEE Vehicle Power and Propulsion Conference, Dearborn, MI, USA, 2009, pp. 304-309, doi: 10.1109/VPPC.2009.5289837.

[25] A 'plug-and-play' electric vehicle battery solution? Inside Webasto's new modular CV battery system. [Online]. Available: https://www.fleetequipmentmag.com/webasto-modular-cv-battery-system/

[26] BQ76940. [Online]. Available: https://www.ti.com/product/BQ76940

[27] Lithium-Ion Battery Ics. [Online]. Available: https://product.minebeamitsumi.com/en/product/category/ics/battery/monitor/

[28] Active Battery Cell Balancing. [Online]. Available: https://www.analog.com/en/technical-articles/active-battery-cell-balancing.html

3/2 power law in magnetoresistance of TmB_4

[1]*Július BAČKAI (3rd year)*
Supervisor: [2]Slavomír GABÁNI

[1]Department of Physics, FEI TU Košice, Slovak Republic
[1,2]Centre of Low Temperature Physics, IEP SAS Košice, Slovak Republic

[1]julius.backai@tuke.sk, [2]gabani@saske.sk

Abstract— **We have studied TmB_4, a strongly anisotropic Ising antiferromagnet with a geometrically frustrated Shastry-Sutherland lattice exhibiting fractional magnetization plateaus, through transport experiments in the presence of a magnetic field. The change of a electrical resistance (R) in response to an external magnetic field (H) provides important information for the characterization of its electronic properties and has found applications in sensor and storage related technologies. In good metals, Boltzmann's theory predicts a quadratic growth in magnetoresistance (MR) at low H and saturation at high fields. On the other hand, a number of nonmagnetic materials with weak electronic correlation and low carrier concentration, such as inhomogeneous conductors, semimetals, narrow gap semiconductors and topological insulators, and two-dimensional electron gas, show positive, non-saturating linear magnetoresistance (LMR). In case of TmB_4, there was shown a non-saturating "3/2" power law magnetoresistance in region of low magnetic field and current directions $I \parallel [110]$ and $I \parallel [100]$, respectively. These results are in strong contrast with work** [1]**, where authors claim tuning magnetoresistance from quadratic to linear.**

***Keywords*— low temperature physics, tetraborides, frustrated systems, magnetoresistance**

I. Introduction

Properties of quantum spins with antiferromagnetic (AF) coupling on frustrated lattices have attracted widespread interest in recent years due to the discovery of a variety new quantum ground states as e.g. spin ice [2], quantum spin liquid-like states [3], and fractional magnetization plateaus on the Shastry-Sutherland lattice [4], among them on insulating $SrCu_2(BO_3)_2$ [5] as well as on the family of metallic and magnetic rare earth (RE) tetraborides, REB_4 [6]–[15]. In insulating $SrCu_2(BO_3)_2$ the exchange interaction is of the Heisenberg type, but on the other hand in metallic REB_4 magnets, the AF exchange interaction between their magnetic moments is of long-range Ruderman-Kittel-Kasuya-Yosida (RKKY) type mediated by conduction electrons. On the other hand, in such systems also the transport properties can be strongly influenced by the magnetic structure. This interplay between charge carriers and magnetism therefore enables to use transport experiments as an indirect probe of the magnetic structures that are present in such systems.

Probably the most investigated REB_4 is thulium tetraboride TmB_4, which orders antiferromagnetically at $T_N = 11.7$ K, has attracted attention for its rich magnetic phase diagram which is strongly biased by crystal field effects at Tm^{3+} ion sites that lift the degeneracy of the $J = 6$ multiplet and lead to a $M_J = \pm 6$ ground state doublet [6], [16]–[19]. The most distinctive features of magnetization M along the c axis are various fractional magnetization plateaus which depend on applied field H. Very recent results concerning the properties of fractional plateaus based mainly on magnetization and heat capacity measurements can be found in [20]–[22].

Nevertheless, in metallic TmB_4, as mentioned above, also conducting electrons can provide additional information about various magnetic states. Such information, based on resistivity ρ, magnetoresistance and Hall effect measurements, were recently obtained in [1], [19], [23]. They show that electronic transport as a function of temperature and magnetic field is a very sensitive probe of scattering processes on the magnetic order / disorder in these frustrated systems.

Here, we present a detailed magneto-resistance measurement in two current direction $I \parallel [110]$ and $I \parallel [100]$. Both directions show same 3/2 growth of magnetoresistance with increasing magnetic field. It seems that temperature does not influence this 3/2 dependence significantly as it was observed in perform measurements at various temperatures.

II. Experimental

The used TmB_4 single crystalline samples were grown by inductive, crucible-free zone melting method, with residual resistivity ratio slightly larger than 14, documenting their high quality. All used samples were cut from one large oriented single crystal. More information about sample preparation can be found e.g. in [24].

Magnetoresistance $\Delta\rho/\rho(\varphi, H) = [\rho(\varphi, H)-\rho(\varphi, H=0)]/\rho(\varphi, H=0)$ measurements were performed using a standard low-frequency ac technique in a commercial PPMS unit equipped with a sample rotation option. The current I was applied along the [110] and subsequently along [100] direction of the Shastry-Sutherland (a-b) plane, whereas those directions were also the axis of rotation. The sample orientation to external magnetic field was changed with a step of $\Delta\varphi = 1°$. In case of axis of rotation along [110] orientation of sample to magnetic field changed from [001] via $[\bar{1}11]$ and $[\bar{1}10]$ to $[00\bar{1}]$, this experimental configuration ensured that magnetic field crossed all crystallographic directions in the sample. In our measurements there was always $H \perp I$, but the field could change its orientation, during sample rotation from

perpendicular to parallel alignment to the Shastry-Sutherland plane (for details see Fig. 1a) and Fig. 1b)).

Since the fractional magnetization plateaus depend on field history [19], [22], the same measuring protocols were used for magnetoresistance as for magnetization measurements.

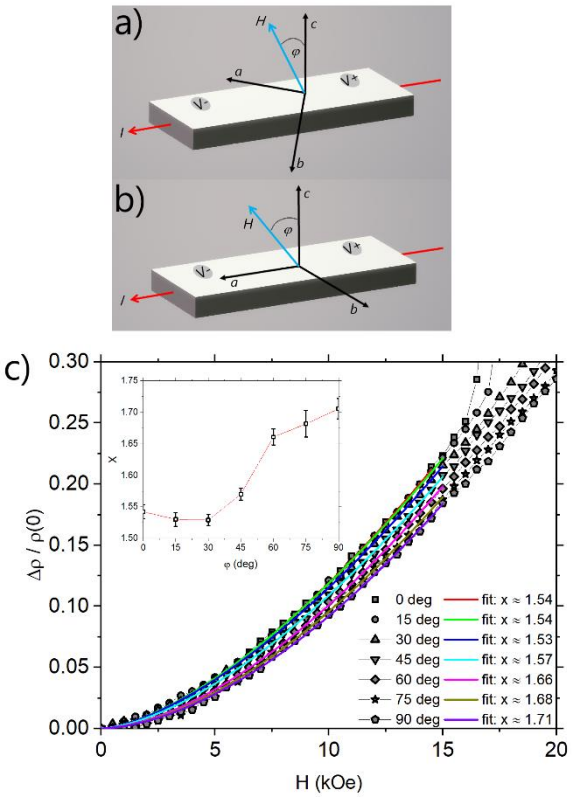


Fig. 1. Layout of magnetoresistance measurements a) current I and axis of rotation along [110] direction, b) current I and axis of rotation along [100] direction. The orientation of applied magnetic field H, which is always perpendicular to I, changed from 0° to 180°. Bottom picture c) shows the evolution of low field dependencies of magnetoresistance $\Delta\rho/\rho(\varphi,H)$ with increasing angle φ between the applied field H and the sample c-axis fitted by $(H/H_0)^x$ expression [1] measured at temperature T = 2 K. This measurement was performed with experimental setup according to Fig. 1a. Inset: the variation of exponent x with angle φ.

III. Results and Discussion

Fig. 1c shows the $\Delta\rho/\rho(\varphi, H)$ dependencies for various angles between $\varphi = 0°$ and $\varphi = 90°$, with current direction along [110]. The analysis of these dependencies shows that $\Delta\rho/\rho(\varphi, H)$ is for all angles positive and in small magnetic fields varies as $\Delta\rho/\rho = (H/H_0)^x$, with x depending on φ (see inset of Fig. 1c). At $\varphi = 0°$ is $x \approx 1.54$, and at $\varphi = 90°$ is $x \approx 1.70$. At Fig. 2 are shown the $\Delta\rho/\rho(\varphi, H)$ dependencies for current flow in [100] direction at two temperatures a) T = 2 K and b) T = 4 K. The obtained results are not in accordance with those in [1] where MR changes from a linear dependence at φ = 0° to a quadratic dependence at φ = 90°, nevertheless our experimental layout was the same. Authors in [1] support their result of linear MR with DFT calculations but they also admit that the calculations are unable to capture the effects of strong correlations precisely. The discrepancy between our results and these obtained in [1] can be explained by the various sample quality (r_R = 38 in [1]). The only theoretical prediction of the 3/2 power law magnetoresistance, we are aware of, has been drawn for two-dimensional electron gas in the presence of small modulated magnetic field with zero average [25], the model which is not in the best match as we did not use a modulated magnetic field. Otherwise, our results are in good agreement with those in [26] and [27], where magnetoresistance of hexaborides SmB_6 and LaB_6 shows approximately $H^{3/2}$ dependence. According to these papers, 3/2 power law in magnetoresistance has origin in presence of carriers in open orbit on the Fermi surface.

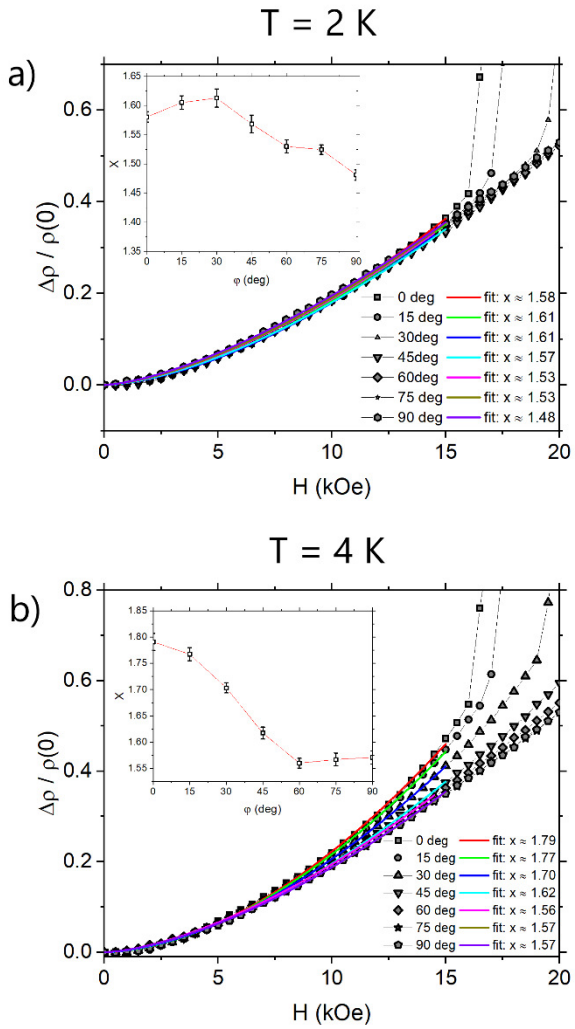


Fig. 2a) and 2b) shows the evolution of low field dependencies of magnetoresistance $\Delta\rho/\rho(\varphi,H)$ with increasing angle φ between the applied field H and the sample c-axis fitted by $(H/H_0)^x$ expression [1] measured at temperature T = 2 K and T = 4 K, respectively. This experimental setup (current along [100] direction according to Fig 1b) was exactly the same as in [1], we have not observed any evolution of exponent x. Insets: the variation of exponent x with angle φ.

IV. Conclusion

In summary we measured dependencies of magnetoresistance of TmB_4 in two different current directions ([110] and [100]). Despite the previously observed quadratic to linear dependence of magnetoresistance with plausible explanation of the origin of linear magnetoresistance in [1], we have not seen any evolution of exponent x in our $\Delta\rho/\rho = (H/H_0)^x$ dependencies. On the other hand, we observed approximately $\Delta\rho/\rho \sim H^{3/2}$ dependence over the entire angle range. However, the difference between these two works is interesting and it should be investigated more detail in the future.

ACKNOWLEDGMENT

This work was supported by projects APVV-17-0020, VEGA 2/0032/20, DAAD-57561069, VA SR ITMS2014+ 313011W856 and by the European Microkelvin Platform. Liquid nitrogen for experiments was sponsored by U.S. Steel Košice, s.r.o.

REFERENCES

[1] S. Mitra *et al.*, "Quadratic to linear magnetoresistance tuning in TmB4," *Phys Rev B*, vol. 99, no. 4, Jan. 2019, doi: 10.1103/PhysRevB.99.045119.

[2] A. P. Ramirez, A. Hayashi, R. J. Cava, R. Siddharthan, and B. S. Shastry, "Zero-point entropy in 'spin ice,'" 1999, Accessed: Feb. 04, 2022. [Online]. Available: www.nature.com

[3] J. S. Helton *et al.*, "Spin Dynamics of the Spin-1/2 Kagome Lattice Antiferromagnet ZnCu3(OH)6Cl2," *Phys Rev Lett*, vol. 98, no. 10, Oct. 2006, doi: 10.1103/PhysRevLett.98.107204.

[4] B. S. Shastry and B. Sutherland, "Exact ground state of a quantum mechanical antiferomagnet" 1981.

[5] H. Kageyama *et al.*, "Exact Dimer Ground State and Quantized Magnetization Plateaus in the Two-Dimensional Spin System," *Phys Rev Lett*, vol. 82, no. 15, p. 3168, Apr. 1999, doi: 10.1103/PhysRevLett.82.3168.

[6] K. Siemensmeyer *et al.*, "Fractional magnetization plateaus and magnetic order in the shastry-sutherland magnet TmB4," *Phys Rev Lett*, vol. 101, no. 17, Oct. 2008, doi: 10.1103/PhysRevLett.101.177201.

[7] G. Will, W. Schäfer, F. Pfeiffer, F. Elf, and J. Etourneau, "Neutron diffraction studies of TbB4 and ErB4," *Journal of The Less-Common Metals*, vol. 82, no. C, pp. 349–355, 1981, doi: 10.1016/0022-5088(81)90238-1.

[8] D. Okuyama *et al.*, "Competition of Magnetic and Quadrupolar Order Parameters in HoB4," *http://dx.doi.org/10.1143/JPSJ.77.044709*, vol. 77, no. 4, Apr. 2008, doi: 10.1143/JPSJ.77.044709.

[9] J. Y. Kim, B. K. Cho, and S. H. Han, "Anisotropic magnetic phase diagrams of HoB4 single crystal," *J Appl Phys*, vol. 105, no. 7, p. 07E116, Feb. 2009, doi: 10.1063/1.3075871.

[10] R. Watanuki *et al.*, "Geometrical Quadrupolar Frustration in DyB4," *J Physical Soc Japan* vol. 74, no. 8, pp. 2169–2172, Nov. 2013, doi: 10.1143/JPSJ.74.2169.

[11] D. Okuyama, T. Matsumura, H. Nakao, and Y. Murakami, "Quadrupolar Frustration in Shastry–Sutherland Lattice of DyB4 Studied by Resonant X-ray Scattering," *J Physical Soc Japan*, vol. 74, no. 9, pp. 2434–2437, Nov. 2013, doi: 10.1143/JPSJ.74.2434.

[12] J. Y. Kim, N. H. Sung, B. Y. Kang, M. S. Kim, B. K. Cho, and J. S. Rhyee, "Magnetic anisotropy and magnon gap state of SmB4 single crystal," *J Appl Phys*, vol. 107, no. 9, p. 09E111, May 2010, doi: 10.1063/1.3365061.

[13] P. Farkašovský, H. Čenčariková, and S. Mat'aš, "Numerical study of magnetization processes in rare-earth tetraborides," *Phys Rev B Condens Matter Mater Phys*, vol. 82, no. 5, p. 054409, Aug. 2010, doi: 10.1103/PHYSREVB.82.054409/FIGURES/6/MEDIUM.

[14] J. Strečka, O. Rojas, T. Verkholyak, and M. L. Lyra, "Magnetization process, bipartite entanglement, and enhanced magnetocaloric effect of the exactly solved spin-1/2 Ising-Heisenberg tetrahedral chain," *Phys Rev E Stat Nonlin Soft Matter Phys*, vol. 89, no. 2, Feb. 2014, doi: 10.1103/PHYSREVE.89.022143.

[15] S. Gabani, K. Flachbart, K. Siemensmeyer, and T. Mori, "Magnetism and superconductivity of rare earth borides," *J Alloys Compd*, vol. 821, p. 153201, Apr. 2020, doi: 10.1016/J.JALLCOM.2019.153201.

[16] F. Iga *et al.*, "Highly anisotropic magnetic phase diagram of a 2-dimensional orthogonal dimer system TmB4," *J Magn Magn Mater*, vol. 310, no. 2 SUPPL. PART 2, Mar. 2007, doi: 10.1016/j.jmmm.2006.10.476.

[17] S. Gabáni *et al.*, "Magnetic structure and phase diagram of TmB4," *Acta Phys Pol A*, vol. 113, no. 1, pp. 227–230, 2008, doi: 10.12693/APHYSPOLA.113.227.

[18] S. Michimura, A. Shigekawa, F. Iga, T. Takabatake, and K. Ohoyama, "Complex magnetic structures of a Shastry-Sutherland lattice TmB 4 studied by powder neutron diffraction analysis," *J Physical Soc Japan*, vol. 78, no. 2, Feb. 2009, doi: 10.1143/JPSJ.78.024707.

[19] S. S. Sunku *et al.*, "Hysteretic magnetoresistance and unconventional anomalous Hall effect in the frustrated magnet TmB4," *Phys Rev B*, vol. 93, no. 17, May 2016, doi: 10.1103/PhysRevB.93.174408.

[20] J. Trinh, S. Mitra, C. Panagopoulos, T. Kong, P. C. Canfield, and A. P. Ramirez, "Degeneracy of the 1/8 Plateau and Antiferromagnetic Phases in the Shastry-Sutherland Magnet TmB4," *Phys Rev Lett*, vol. 121, no. 16, Oct. 2018, doi: 10.1103/PhysRevLett.121.167203.

[21] D. Lançon *et al.*, "Evolution of field-induced metastable phases in the Shastry-Sutherland lattice magnet TmB4," *PhRvB*, vol. 102, no. 6, p. 060407, Aug. 2020, doi: 10.1103/PHYSREVB.102.060407.

[22] M. Orendáč *et al.*, "Ground state and stability of the fractional plateau phase in metallic Shastry–Sutherland system TmB4," *Sci Rep*, vol. 11, no. 1, p. 6835, Dec. 2021, doi: 10.1038/S41598-021-86353-5.

[23] L. Ye, T. Suzuki, and J. G. Checkelsky, "Electronic transport on the Shastry-Sutherland lattice in Ising-type rare-earth tetraborides," *Phys Rev B*, vol. 95, no. 17, May 2017, doi: 10.1103/PhysRevB.95.174405.

[24] N. Shitsevalova, "Crystal Chemistry and Crystal Growth of Rare-Earth Borides," *Rare-Earth Borides*, pp. 1–243, Oct. 2021, doi: 10.1201/9781003146483-1.

[25] A. Matulis and F. M. Peeters, "Magnetoresistance of a two-dimensional electron gas in weakly modulated magnetic fields," *Phys Rev B*, vol. 62, no. 1, 2000.

[26] J. C. Cooley, M.C. Aronson, A. Lacerda, Z. Fisk, P. C. Canfield, and R. P. Guertin, "High magnetic fields and correlation gap in SmB6," *Phys Rev B*, vol. 52, no. 10, 1995.

[27] A. J. Arko *et al.*, "de Haas-van Alyhen effect and the Fermi surface of LaBs," *Phys Rev B*, vol. 13, p. 15, 1976.

Effect of increased disorder on the pair-breaking mechanism in Mo_2N thin films

[1]*Marek KUZMIAK (4th year)*
Supervisor: [2]Pavol SZABÓ

[1]Dept. of Physics, FEI TU of Košice, Slovak Republic
[1,2]Centre of Low Temperature Physics, Institute of Experimental Physics of Košice, Slovak Academy of Sciences, Slovak Republic

[1]marek.kuzmiak@tuke.sk, [2]pszabo@saske.sk

***Abstract*—My thesis is focused on the experimental study of the mechanism of the superconductor-insulator transition (SIT) in strongly disordered superconducting thin film systems. In this article, the physical properties of highly disordered superconducting ultrathin γ-Mo_2N films are studied by applying magnetotransport measurements. We show, that SIT in this system occurs at thicknesses between 2-3 nm. The magnetic field vs. temperature (B-T) phase diagrams of our samples, constructed from the transport measurements show, that with the increasing disorder, the mechanism of the magnetic suppression of superconductivity is changing. The results, presented in this paper serve as supplementary material for a forthcoming review article on SIT in thin γ-Mo_2N films and will be an important part of my thesis.**

***Keywords*—superconductivity, molybdenum nitride (Mo_2N), orbital pair-breaking effect, spin-paramagnetic pair-breaking effect, magnetotransport.**

I. Introduction

The superconductor-insulator transition (SIT) in two dimensions is a continuous quantum phase transition in the limit of low temperatures driven by external parameters like disorder, magnetic field, or carrier concentration [1]. Two general mechanisms of SIT are known. These mechanisms can modify the superconducting wave function $\psi(\vec{r}) = \Delta e^{i\Theta(\vec{r})}$ either by suppressing the amplitude Δ – fermionic scenario [2] or by breaking its phase coherence Θ – bosonic scenario [3]. Our previous articles have shown that increased disorder in γ-Mo_2N ultrathin films leads to the fermionic scenario of the suppression of superconductivity [4,5].

It is well known, that the application of an external magnetic field will suppress the superconducting properties [6]. The magnetic field can break Cooper pairs of the superconducting condensate in two different ways. The first mechanism is the orbital pair-breaking effect, which is explained by the suppression of superconductivity via the formation of Abrikosov vortices. It's typical for this mechanism, that the B-T phase diagram (temperature dependence of the upper critical magnetic field $B_{C2}(T)$) has a linear character at temperatures $T \to T_C$. The second mechanism is the spin-paramagnetic pair-breaking effect, which is explained as the effect of the magnetic field on the spin of the paired electrons of the Cooper pairs. At this mechanism, the B-T phase diagram has a typical square root dependence $\propto \sqrt{T_C - T}$ at $T \to T_C$ [7].

The present paper deals with the transport measurements and the effect of increasing magnetic field on the transport properties of the investigated γ-Mo_2N thin films. We show, that the mechanism of the magnetic pair-breaking in disordered γ-Mo_2N thin films depends on the degree of disorder. While in weakly/mediately disordered thin films superconductivity is suppressed with orbital pair-breaking, we identified the presence of a paramagnetic pair-breaking in the ultrathin film that was close to critical disorder.

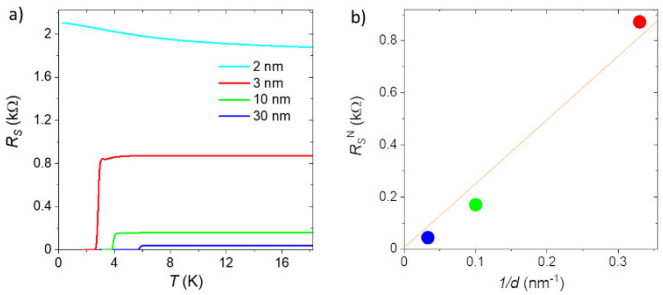


Fig. 1. **Temperature dependence of sheet resistance $R_S(T)$ in a zero magnetic field.** Transport measurements (a) were performed on 2 nm (cyan), 3 nm (red), 10 nm (green), and 30 nm (blue) samples. The thickness dependence of the normal state sheet resistance of the superconducting samples R_S^N in the reciprocal $1/d$ scale is shown in (b). The solid orange line is a guide for the eyes.

II. Experiment

Our Mo_2N thin films of 2 nm, 3 nm, 10 nm, and 30 nm thicknesses were prepared by reactive magnetron sputtering onto a sapphire substrate from a Mo target in an argon-nitrogen mixture gas atmosphere. The subsequent crystallographic structure of our samples was characterized by X-ray diffraction. It was detected that our films are formed by a stoichiometric γ-Mo_2N phase with a typical NaCl structure [8].

The magnetotransport measurements were realized in a Quantum Design Physical Properties Measurement System (PPMS) using transport and resistivity options in the temperature range from 300 K to 0.4 K. This device allows for measurements in magnetic fields up to 9T. First, the sample was attached with vacuum grease on a standard Quantum Design Helium-3 Puck and DC Resistivity/ETO Sample Puck for resistivity measurements. These Pucks represent a simple and effective way to secure and wire samples for electrical measurements. Then contact connections were made using

50 µm silver wires that were fixed to samples via silver paint. These were attached to the samples in a four-probe Van der Pauw configuration [9]. During the measurement of our samples, the lowest possible current from the PPMS measurement mode of 0.01 mA at a frequency of 30 Hz was applied. Subsequently, measurements of temperature dependencies at different external magnetic fields were carried out on our γ-Mo_2N samples. The transport measurements of the sheet resistance $R_S(T)$ were performed using the Van der Pauw four-probe technique. The magnetic field was applied perpendicular to the plane of the γ-Mo_2N thin films.

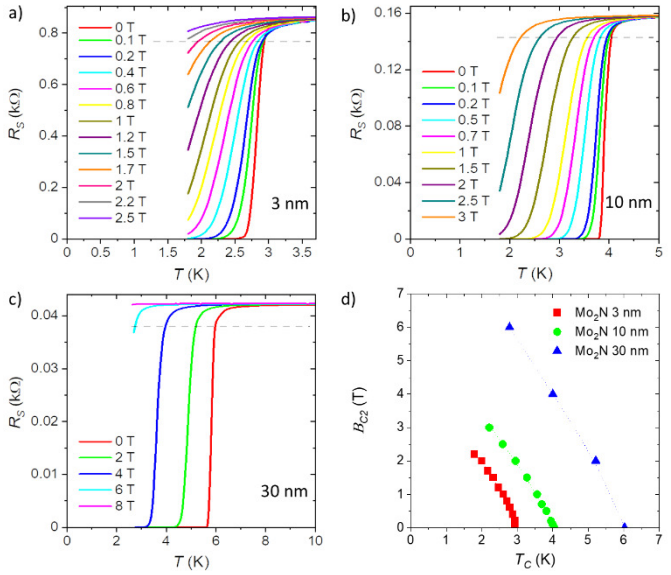


Fig. 2. **Temperature dependencies of sheet resistance $R_S(T)$ in different magnetic fields.** Magnetotransport measurements were performed on 3 nm (a), 10 nm (b), and 30 nm (c) samples. Fig. (d) shows B-T phase diagrams of the thin films with thicknesses of 3 nm (red symbols), 10 nm (green symbols), and 30 nm (blue symbols). The normal state transitions were determined from magnetotransport data at 90 % of the normal state sheet resistance $R_S(T)$ (grey dashed line in a-c).

III. Results and Discussion

Figure 1 (a) shows data, which were obtained from transport measurements of sheet resistance $R_S(T)$ in a zero magnetic field. These measurements were performed on 2 nm, 3 nm, 10 nm, and 30 nm γ-Mo_2N samples. As can be seen, with reducing thin film thickness, the normal state sheet resistance increases and the critical temperature T_C is suppressed. Our samples reveal superconductivity down to 3 nm thickness. The $R_S(T)$ curve of the 2 nm thin film already shows an insulating character, which indicates that the SIT occurs somewhere at thicknesses between 3 nm and 2 nm. This is in agreement with our previous studies [5], but also with the observations of other groups [10,11]. Figure 1 (b) shows the thickness d dependence of the normal state sheet resistance R_S^N of the studied samples on a $1/d$ scale. The $R_S^N(1/d)$ dependence of the disordered films is often used for basic characterization of the quality of a series of thin films with different thicknesses. Good quality continuous thin films reveal linear suppression of R_S^N in the reciprocal thickness scale [10]. The quasi-linear dependence of $R_S^N(1/d)$ of our thin films confirms the high quality of our samples.

Our magnetotransport data are shown in Figure 2. Temperature dependencies of the sheet resistance $R_S(T)$ of the 3 nm, 10 nm, and 30 nm samples measured at different fixed magnetic fields are shown in panels (a), (b), and (c), resp. Samples with 3 nm and 10 nm thicknesses were measured up to B = 3 T and the 30 nm thin film to B = 8 T magnetic fields. It can be seen that as the value of the applied magnetic field increases, the sheet resistance remains constant in the normal state and the superconducting transition shifts to lower temperatures in all our samples. It's the typical behaviour of superconductors in the presence of a magnetic field.

Figure 2 (d) shows the B-T phase diagram of the studied samples. The T_C values were determined at 90 % of the normal state sheet resistances of the $R_S(T)$ curves measured at different fixed magnetic fields, which are marked with grey dashed lines in Figure 2 (a-c). The determined $B(T_C)$ dependencies are shown in Figure 2 (d) with red (3 nm), green (10 nm), and blue (30 nm) symbols. Although our B-T phase diagrams show the B_{C2} values only from the proximity of T_C it is obvious, that with increasing disorder not only the values of T_C and B_{C2} are changed, but also the curvature of the $B_{C2}(T)$ curves is changing. The B_{C2}(T) values determined on the 30 nm and 10 nm thin films show linear temperature dependencies near T_C, which is typical for the orbital pair-breaking mechanism. On the other hand, the critical field of the 3 nm thin film reveals a negative curvature with a square root dependence $\propto \sqrt{T_C - T}$. Such a curvature is typical for the paramagnetic critical field B^P [12], where the magnetic pair breaking is realized through Zeeman spin-splitting effects. These results point to the apparent change in the pair-breaking mechanism due to increasing disorder in γ-Mo_2N thin films.

IV. Conclusion

This paper is dealing with the experimental study of transport properties in the series of γ-Mo_2N thin films with thicknesses 2, 3, 10, and 30 nm. We have shown, that the reduced thickness causes an increase in disorder, leading to a superconductor-insulator transition at thin film thicknesses between 3 nm and 2 nm.

The B-T phase diagrams of our samples show that with increasing disorder not only the values of B_{C2} and T_C are suppressed, but also the mechanism of the magnetic pair breaking changes. The orbital pair breaking effect, observed on samples with the lower disorder, changes to the paramagnetic effect in close proximity to the superconductor-insulator transition. These surprising results will be further studied using local spectroscopic measurements in our following papers.

Acknowledgement

This work was supported by the Ministry of Education, Science, Research and Sport of the Slovak Republic and the Slovak Academy of Sciences under contracts No. VEGA 1/0743/19, VEGA 2/0058/20, APVV-18-0358, APVV-20-0425, Doktogrant APP0292, the European Microkelvin Platform, the COST action CA21144, and by the U.S. Steel Košice.

References

[1] F. Wang, J. Biscaras, A. Erb, and A. Shukla, *Superconductor-Insulator Transition in Space Charge Doped One Unit Cell*

Bi2.1Sr1.9CaCu2O8+x, Nature Communications 2021 12:1 **12**, 1 (2021).

[2] A. M. Finkel'stein, *Metal-Insulator Transition in Disordered Two-Dimensional Electron Systems*, Sov. Phys. JETP **59**, 212 (1984).

[3] M. P. A. Fisher, *Quantum Phase Transitions in Disordered Two-Dimensional Superconductors*, Phys. Rev. Lett. **65**, 923 (1990).

[4] M. Kuzmiak, M. Kopčík, F. Košuth, V. Vaňo, P. Szabó, V. Latyshev, V. Komanický, and P. Samuely, *Suppressed Superconductivity in Ultrathin Mo2N Films Due to Pair-Breaking at the Interface*, J Supercond Nov Magn **35**, 1775 (2022).

[5] M. Kuzmiak and at al., *To Be Published*, (n.d.).

[6] M. Žemlička, M. Kopčík, P. Szabó, T. Samuely, J. Kačmarčík, P. Neilinger, M. Grajcar, and P. Samuely, *Zeeman-Driven Superconductor-Insulator Transition in Strongly Disordered MoC Films: Scanning Tunneling Microscopy and Transport Studies in a Transverse Magnetic Field*, Phys Rev B **102**, 180508 (2020).

[7] S. Mozaffari et al., *Superconducting Phase Diagram of H3S under High Magnetic Fields*, Nature Communications 2019 10:1 **10**, 1 (2019).

[8] M. Kozejova, V. Latyshev, V. Kavecansky, H. You, S. Vorobiov, A. Kovalcikova, and V. Komanicky, *Evaluation of Hydrogen Evolution Reaction Activity of Molybdenum Nitride Thin Films on Their Nitrogen Content*, Electrochim Acta **315**, 9 (2019).

[9] L. J. van der Pauw, *A Method of Measuring Specific Resistivity and Hall Effect of Discs of Arbitrary Shape*, Philips Res. Rep **13**, 1 (1958).

[10] T. Tsuneoka, K. Makise, S. Maeda, B. Shinozaki, and F. Ichikawa, *Localization and Pair Breaking Parameter in Superconducting Molybdenum Nitride Thin Films*, Journal of Physics: Condensed Matter **29**, 015701 (2016).

[11] K. Makise, F. Ichikawa, T. Asano, and B. Shinozaki, *Field-Tuned Superconductor–Insulator Transitions and Hall Resistance in Thin Polycrystalline MoN Films*, Journal of Physics: Condensed Matter **30**, 065402 (2018).

[12] Y. Matsuda and H. Shimahara, *Fulde-Ferrell-Larkin-Ovchinnikov State in Heavy Fermion Superconductors*, J Physical Soc Japan **76**, 051005 (2007).

Study of thermoplastic starch-based materials

[1]*Leoš ONDRIŠ (1st year)*
Supervisor: [2]*Oľga FRIČOVÁ*

[1,2]Department of Physics, FEI TUKE, Slovak Republic

[1]leos.ondris@tuke.sk, [2]olga.fricova@tuke.sk

***Abstract*— Due to the high consumption of plastics and their adverse effects on the environment, new materials based on biodegradable polymers are being investigated. Starch is a versatile natural polymer material used mainly in the food industry. Modification plays an important role in the application of starch in other fields. The most common modifications are plasticization, cross-linking, mixing with other polymers and reinforcing starch with (nano)fillers to create (nano)composite materials. The physical and mechanical properties of starch-based materials depend on their structure and molecular mobility, which can be studied using X-ray diffractometry, dynamic-mechanical thermal analysis and nuclear magnetic resonance.**

***Keywords*— cross-linked starch, dynamic-mechanical analysis, nanocomposites, nuclear magnetic resonance, thermoplastic starch, X-ray diffractometry**

I. Introduction

Polymer materials are an invaluable part of our daily lives. The largest representative of these materials are plastics, which are used in various branches of industry. The massive production of plastics is also associated with a large impact on the environment, particularly in terms of plastic waste and lack of recycling. Another problem is that plastics are mostly obtained from fossil (non-renewable) resources. The solution to these problems lies in replacing conventional synthetic plastic materials with biodegradable ones that can be obtained from renewable sources like cellulose, chitin, chitosan, starch, and others [1][2]. Native starch has found wide use in the food industry. In order to use starch in other branches of the industry, it has to be subject to thermomechanical processing together with small molecular weight compounds (called plasticizers). The resulting material is starch with thermoplastic properties (TPS) [3]. TPS has some unsatisfactory properties in wet and dry environments like high viscosity, and insufficient thermal stability. The main disadvantage of TPS is its retrogradation which causes embrittlement of the material. To eliminate these drawbacks and to improve (selected) TPS properties, TPS with cross-linked chains, its blends with other polymers and (nano)composite TPS-based materials can be prepared [3][4].

II. Starch

Starch is a natural renewable polymer material that belongs to polysaccharides. Starch is produced in higher plants in amyloplasts (by photosynthesis) and serves as a storage substance accumulating in the storage organs of plants in the form of grains. Among the most commercially used sources of starch are potato, rice, wheat, maize, cassava, banana, and others [1][2]. Starch grains can be found in seeds, tubers, roots, leaves, fruits, stems and pollen of plants. Depending on the plant species, starch grains occur in different sizes (0.1-100 μm) and morphologies (polyhedral, elliptical, semi-spherical, spherical, irregular). Different morphologies of starch grains can also be observed in different parts of the same plant [3][4][5].

Starch grains are heterogeneously composed of two main different polysaccharides – primarily linear amylose and highly branched amylopectin (Fig. 1) which are formed by glucose units connected to each other by α-D-(1,4) and α-D-(1,6) glycosidic bonds [3][5]. Other (minor) components of the starch grains are water, proteins, lipids, different forms of phosphorus and other substances which can interact with polysaccharides [2][4]. The percentage of amylopectin in the starch grain is around 65-85% depending on the source. The subscript *n* on Fig. 1 is known as degree of polymerization (DP) and specifies the length of polymer chain. Red numbers in Fig. 1 represent labelling of carbons in the glucose unit.

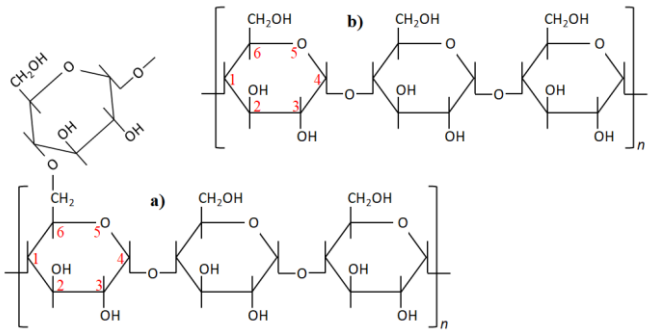


Fig. 1. Chemical structure of amylopectin a) and amylose b)

Starch extracted from starch grains is semicrystalline. The percentage of crystalline phase is approximately 20-45% given by the starch source. The short branching chains of amylopectin are mainly responsible for crystallinity in starch grains. On the other hand, branching points of amylopectin and linear amylose chains form amorphous regions [1][3].

A. Amylopectin

Amylopectin is a highly branched macromolecule, α-D-(1,6) glycosidic bonds are amylopectin branching points (Fig. 1a). The molar mass M of amylopectin chains is 10^7-10^8 g/mol [1][2][3].

Parallel-arranged amylopectin chains form a double-helices that are arranged in crystalline regions of the A-, B-, or C-type. The disordered (amorphous) regions contain branching points. Amylopectin chains may be intertwined with amylose chains [3][5][6].

B. Amylose

Amylose chains are predominantly linear formed by α-D-(1,4) glycosidic linkages (Fig. 1b). The molar mass M of amylose is 10^5-10^6 g/mol. Some amylose chains may contain side chains. Most starches contain a mixture of both linear and branched amylose chains [2][5].

The amylose chains form left-handed single-helical structures. In this state, amylose is unstable. The amylose chains can interact with minor components in starch grains to form inclusion complexes that are arranged in V-type and E_H-type crystalline regions. In some types of starches, the linear chains of amylose can group together and create double-helical structures (thus increasing the stability of the system) and both A- and B-type crystalline structures can be formed [3][5].

III. Thermoplastic starch

Starch is widely available with minimal extraction costs. Due to multiple H-bonds between chains (strong inter- and intra- molecular forces) native starch does not behave as a thermoplastic and it cannot be melted as it degrades before melting [1][3]. For this reason, native starch has to be modified. Chemical modification includes various methods like hydrolysis, etherization, esterification or oxidation [7]. Thermomechanical processing involves various physical and chemical processes (water diffusion, expansion of starch grains, gelatinization, and crystallization). In terms of phase transitions, plasticization/gelatinization is very important because it is involved in the conversion of native starch into a thermoplastic material [3][4].

The process of plasticization is quite complex, as the starch undergoes significant structural changes and thus also changes in its properties. The preparation of TPS includes the following steps: the starch grains are mixed with water or with an aqueous solution of a low molecular weight compounds called plasticizers (glycerin, urea, formamide, sorbitol, …) at elevated temperature (usually around 60-120°C). This swells the starch grains and also disrupts the ordered structures of the amylopectin chains - a gel is formed that is cooled down immediately. In the last step, the sample is dried to remove excess water from the material [1][8].

TPS is composed of a destructured (gelatinized) starch containing one or more plasticizers in his structure. TPS behaves like a thermoplastic and (as well as native starch) is completely biodegradable [3][9]. An important property of starch-based materials is the glass transition temperature T_g. The glass transition is a process in amorphous regions of a polymer in which the polymer chains gradually undergo transition from a glassy state to a rubbery (highly elastic) state as the temperature increases (the process is reversible), due to the changes in their molecular mobility. In the glassy state (below T_g), the polymer is brittle, hard (behaves like glass), the chain segments are frozen in their conformations. In the rubbery state (above T_g) the high molar mass polymer is elastic (the low molar mass polymer is liquid) because the chain segments can move relatively freely [1][3].

TPS is a hydrophilic and flexible material. His flexibility depends on the type (source) of starch and the amount and type of plasticizers. As the plasticizer content increases, the flexibility of the material can increase as it is e.g. for glycerol, when the starch chains and their parts can move more freely, and the T_g decreases [1][3][9].

The plasticizer has the ability to form starch-plasticizer hydrogen (H-) bonds instead of bonds between starch chains, resulting in release of chain segmental motion. At higher plasticizer content, some plasticizer molecules are not bound to the starch chains, thus domains rich in plasticizer and in starch are present in TPS bulk (Fig. 2). In the plasticizer-rich domains, the mobility of the starch chains (segments) is determined by the mobility of the plasticizer molecules and the glass transition is observed at lower temperature than in the starch-rich domains [10].

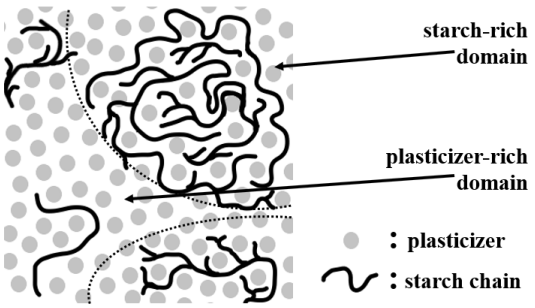


Fig. 2 Inhomogeneous structure of starch-plasticizer mixture

TPS has several drawbacks – mainly insufficient mechanical properties and an unstable structure (especially in humid environments). When stored at temperatures above its glass transition, crystallization of amylose chains and rearrangement of amylopectin chains into ordered structures occur, i.e. the proportion of crystalline phase in TPS increases and the material becomes brittle [9].

IV. Cross-linked starch

Another way to tune the starch properties is chemical treatment to cross-link native starch chains. Many compounds can be use as cross-linking agents, such as carbodiimides, tannic acid, aldehydes, epichlorohydrin, aldehydes, and many others [7][11]. Cross-linking of starch is possible in the presence of chemical agents capable of forming ester or ether linkages between hydroxyl (–OH) groups. During starch cross-linking, new covalent bonds are formed between chains acting as intra- and inter- molecular bridges. Cross-linking causes a change in functional properties, water resistance, mechanical strength (limited mobility of starch chains) and thermal stability of starch. Starch cross-linking increases the degree of polymerization and consequently starch molecular mass [12][13]. Oxidation is a widely used chemical modification that leads to the cross-linking, as in the case of dialdehyde starch (DAS) [7][11].

Native starch can be oxidized to predetermined dialdehyde content, thereby modifying its properties in a targeted manner [11][14][15]. DAS is mainly used in the paper industry as wet strength improver. However, due to intrinsic toxicity of the cross-linked agents, use of the DAS in foodstuffs is very risky and it is not permitted. Nevertheless, DAS has great potential as a component of biodegradable plastics for packaging purposes [11][14].

If the DAS is mixed with, for example, (at least) 25% of glycerol and then heat-treated, thermoplastic dialdehyde starch (TPDAS) with thermoplastic properties and a cross-linked structure heat-processable into thin films is produced [7][11]. Some DAS with specific cross-linking agents (disemicarbazone, dihydrazone, etc.) form stable complexes with heavy metals with a possibility of use as biodegradable traps for heavy metal ions from sewage and soil [14].

TPDAS shows longer biodegradation process than TPS but its resistance to starch re-crystallization during storage is higher. TPDAS can be applied in the medical, drug release, packaging, and agricultural fields [7][15]. To extend the use of these materials as biodegradable for various biotechnological applications, further research monitoring the effect of microorganisms on their biodegradation process is important [7].

V. TPS NANOCOMPOSITES

A composite material (composite) is composed of two different components referred to as matrix and filler (reinforcement). The matrix is a continuous component of the composite, its task is to bind and protect the filler that is dispersed in it. The properties of a given composite material are influenced by the chemical and physical properties of the matrix and filler, the amount and shape of the filler, matrix/filler and filler/filler interaction, and the type of preparation process. In most cases, the final product (composite) has different properties than the substances from which it was made up. Polymer composites, sometimes referred to as reinforced plastics, are the most widely used in materials engineering. Compared to other composites, they are characterized by corrosion resistance, low density, ease of preparation and resistance to external influences. They have high specific strength and specific modulus. Most of these composites are also recyclable [16][17][18].

Polymer nanocomposites are composites consisting of a polymer matrix and a nanofiller whose particles have (at least) one dimension smaller than 100 nm. Currently, they are one of the most advanced and prospective materials in terms of research. The disadvantage of polymer (nano)composites is their inferior heat resistance [18][19].

The most widely used nanofiller for polymer matrices is a clay called montmorillonite (MMT). Montmorillonite belongs to the group of smectites, i.e. highly expandable silicate clays, sometimes referred to as phyllosilicates or layered silicates. Each layer is composed of two tetrahedral sheets with an octahedral sheet between them. The large value of the active surface area together with the short distances between the nanofiller particles dispersed in the polymer matrix cause significant changes in the properties of the nanocomposites at very low nanofiller concentrations (<3-5% wt.) [18][20]. Phyllosilicates allow introducing surrounding molecules (including polymer chains) into their interlayer space resulting in an intercalated structure. Individual layers of phyllosilicates are separated in exfoliated nanocomposites [20].

Several scientific studies have shown that the reinforcement of TPS with nanofillers improves its water and water vapour resistance, and some TPS mechanical properties. However, all properties of TPS nanocomposite depend on both the amount and the distribution of nanofiller in the TPS matrix [1][3][4][21].

VI. POLYMER BLENDS WITH TPS

Preparing of blends of TPS with other (biodegradable) polymers is an important modification of TPS. It is very useful, easy to prepare, and low-cost technique. By blending TPS with different polymers it is possible to obtain a material with different properties (mechanical, thermal, water-resistant, rheological, etc.) which increases the range of TPS use. Most blend components are not thermodynamically compatible and therefore they form a multiphase system. In TPS blends, starch can be a continuous or dispersed phase according to its amount in the blend [1][3][4][22].

VII. TECHNIQUES FOR STUDY OF TPS-BASED MATERIALS

Experimental techniques used to characterize properties and structure of starch-based materials include X-ray powder diffractometry, small angle neutron scattering, dynamic-mechanical analysis, differential scanning calorimetry, thermogravimetric analysis, nuclear magnetic resonance, infrared, Raman spectroscopies, transmission and scanning electron microscopies, and rheology [23].

A. X-ray diffractometry

X-ray diffractometry (XRD) is an inexpensive and non-destructive method that is used to detect crystalline phases in material under study. The principle of XRD is the interaction of X-rays with atoms or ions forming the crystallographic lattice of the material under study. The diffraction process on polycrystalline samples is described by Bragg's law:

$$2d_{\mathrm{hkl}} \sin\vartheta = n\lambda, \tag{1}$$

where d_{hkl} is the interplanar distance of two crystallographic planes, hkl are the Miller indices, ϑ is the angle that the incident X-ray beam makes with the sample plane (diffraction angle), n is an integer (the order of diffraction), and λ is the wavelength of the incident (ideally) monochromatic (radiation of the same wavelength) X-rays. Obtained diffractograms, i.e. the dependences of the intensity of diffracted X-ray on the incidence angle ϑ. Diffracted X-rays carry information about structure of the material [24][25][26]. In the diffractograms of starch-based materials, a broad amorphous maximum corresponding to the amorphous phase can be observed which is superimposed with diffraction peaks at particular angles ϑ, according to which the type of crystalline structure of starch can be identified [3][26].

B. Dynamical-mechanical analysis

The dynamic-mechanical thermal analysis (DMA) method examines the response of a material to an applied external harmonic force as a function of temperature. Based on these measurements, we can assess mechanical properties of materials in the relevant temperature ranges. If we write the applied mechanical stress in the form:

$$\sigma(t) = \sigma_0 \sin(\omega t), \tag{2}$$

then for the resulting strain lagging behind the stress we get the equation:

$$\gamma(t) = \gamma_0 \sin(\omega t - \delta), \tag{3}$$

where σ_0 and γ_0 have the meaning of stress and strain amplitude, respectively, ω is the angular frequency of applied harmonic stress, and t is time. Equation (3) describes the harmonic deformation of a viscoelastic material (a material that combines elastic and viscous behavior characterized by a storage (E') and loss (E'') moduli, respectively), for example polymeric materials. The parameter δ is the phase shift, also called the loss angle. In the case of polymers, various relaxation transitions associated with changes in the molecular motion can be investigated, such as the glass transition, secondary transitions, melting, the effect of additives or structural changes during storage. Relaxation processes are manifested in the temperature dependences of $\tan\delta$ and loss modulus as local maxima, storage modulus decreases with increasing temperature [27][28]. This information is important for appropriate use of the studied material.

C. Nuclear magnetic resonance

Nuclear magnetic resonance (NMR) is nondestructive spectroscopic method to investigate the structure and molecular dynamics/mobility of starch-based materials. NMR is based on the interaction of a radio-frequency (MHz) electromagnetic field (rf-EMF) with the magnetic moments (related to spins) of the nuclei of atoms, which are placed in a strong external magnetic field B_0 [25][26]. A nucleus with nuclear spin I (I≠0) in a magnetic field B_0 takes on 2I+1 orientations with different energies. A given nucleus interacts also with the local magnetic fields generated by the adjacent nuclei. These local magnetic fields are unique to the structure of the material. By irradiation the sample (immersed in the magnetic field B_0) with rf-EMF, the nuclei absorb energy changing the orientation of the nuclear spins from a low-energy to high-energy state (excited state) and the magnetization (total magnetic moment) of the sample changes. This absorption of energy, called resonance, occurs only if the resonance condition is met:

$$|\Delta E| = |\gamma \hbar B_0|, \tag{4}$$

where ΔE is the difference in the energy of two states with different orientation of magnetic moment in the B_0 field, γ is the gyromagnetic ratio and ħ is the reduced Planck constant. When the rf-EMF is removed, the nuclei relax back to their low-energy state, the sample magnetization changes; it is detected as an induced signal so called free induction decay (in time domain) which is recorded. This signal is then converted via Fourier transformation into frequency domain – NMR spectrum, which provides information about the structure and molecular dynamics/motion of the samples [26][29].

^{1}H NMR experiments are frequently used for characterizing molecular dynamics in the studied samples. Proton NMR spectra of solids are usually broad and in the case of rigid structure poorly resolved due to the strong dipole-dipole interactions between ^{1}H nuclear spins which are not averaged by molecular motion, the widths of NMR signals narrows with increasing molecular mobility. Using ^{13}C NMR measurements, it is possible to observe the changes in morphology in TPS-based materials in detail, because the shape of the resonance signal of carbon C1 in starch structure (Fig. 1) is very sensitive to the ordering of starch chains in the scale smaller than nm, which cannot be detected by XRD [8].

My research is focused on the characterization of thermoplastic starch and thermoplastic starch-based biodegradable polymer blends. These materials will be studied using above mentioned experimental techniques, which can provide comprehensive information about their structure and changes of its properties over time.

References

[1] S. Thomas, P. M. Visakh, A. P. Mathew, "Advances in natural polymers: Composites and nanocomposites: Starch based blends, composites and nanocomposites," in *Advanced structure materials*, vol. 18, Springer, Berlin, Heidelberg, 2013, pp. 121-154.

[2] E. Bertoft, "Understanding starch structure: Recent progress," in *Agronomy*, vol. 7, 2017, pp. 29.

[3] M. N. Belgacem, A. Gandini, *Monomers polymers and composites from renewable resources: Starch: Major sources, properties, and applications as thermoplastic materials*, Kidlington: Elsevier Ltd., 2008, ch. 15.

[4] P. J. Halley, L. Avérous *Starch polymers: From genetic engineering to green applications: Starch polymers: From the field to industrial products*, Elsevier Ltd. 2014, ch. 1.

[5] S. Peréz, E. Bertoft, "The molecular structure of starch components and their contribution to the architecture of starch granules: A comprehensive review," in *Starch/Stärke*, vol 62, 2010, pp. 389-420.

[6] A. Buléon, G. Véronèse, J.-L. Putaux, "Self-association and crystallization of amylose," in *Australian Journal of Chemistry*, vol 60, 2007, pp. 706-718.

[7] Y.-L. Du, et al. "Biodegradation behaviors of thermoplastic starch (TPS) and thermoplastic dialdehyde starch (TPDAS) under controlled composing conditions," in *Polymer Testing*, vol 27, 2008, pp. 924-930.

[8] A. Šoltýs, V. Hronský, N. Šmidová, D. Olčák, F. Ivanič, I. Chodák. "Solid -state ^{1}H and ^{13}C NMR of corn starch plasticized with glycerol and urea," in *European Polymer Journal*, vol 117, 2019, pp. 19-27.

[9] A. M. Nafchi, M. Moradpour, M. Saeidi, A. K. Alias, "Thermoplastic starches: Properties, challenges and prospects," in *Starch/Stärke*, vol 65, 2013, pp. 61-72.

[10] K. Kawai, Y. Hagura, "Discontinuous and heterogenous glass transition behavior of carbohydrate polymer-plasticizer system," in *Carbohydrate Polymers*, vol. 89, 2012, pp. 839-841.

[11] H. Peidayesh, Z. Ahmadi, H. A. Khonakdar, M. Abdouss, I. Chodák, "Fabrication and properties of thermoplastic starch/montmorillonite composite using dialdehyde starch as a crosslinker," in *Polymer International*, vol 69, 2019, pp. 317-327.

[12] A. Korkut, K. Kahraman, "Production of cross-linked resistant starch from tapioca starch and effect of reaction conditions on the functional properties, morphology, X-ray pattern, FT-IR spectra and digestibility," in *Journal of Food Measurement and Characterization*, vol 15, 2021, pp. 1693-1702.

[13] A. G. Gerezgiher, T. Szabó, "Crosslinking of starch using citric acid," in *Journal of Physics: Conference Series*, vol 2315, 2022, 012036.

[14] M. Fiedorowicz, A. Para, "Structural and molecular properties of dialdehyde starch," in *Carbohydrate Polymers*, vol 63, 2006, pp. 360-366.

[15] J. Yu, P. R. Chang, X. Ma, "The preparation and properties of dialdehyde starch and thermoplastic dialdehyde starch," in *Carbohydrate Polymers*, vol 79, 2010, pp. 296-300.

[16] T.-D. Ngo, *Composite and nanocomposite materials – from knowledge to industrial applications: Introduction to composite materials.* Intechopen, 2020, ch. 1.

[17] V. V. Vasiliev, E. V. Morozov, *Advanced mechanics of composite materials and structural elements*, 3rd. ed., Elsevier Ltd. 2013, ch. 1.

[18] J. Krizbergs, V. Gutakovskis, *Composite materials*, [online] 2016, Available: <https://www.gzs.si/LinkClick.aspx?fileticket=Shxm3khAhiA%3D&portalid=129>.

[19] S. Vyazovskin, N. Koga, Ch. Schick, "Recent advances, techniques and applications: Polymer nanocomposites," in *Handbook of thermal analysis and calorimetry*, vol 6, Elsevier, 2018, ch. 11.

[20] M. Zoveidavianpoor, *Current topics in the utilization of clay in industrial and medical applications: Montmorillonite: An introduction to properties and utilization,* IntechOpen, 2018, ch. 1.

[21] A. Baran et al, "Effect of urea and glycerol mixture on morphology and molecular mobility in thermoplastic starch/montmorillonite-type nanofiller composites studied using XRD and NMR," in *Journal of Polymer Research*, vol 29, 2022, pp. 257.

[22] O. Fričová, M. Hutníková, M. Kovaľaková, A. Baran, "Influence of aging on molecular motion in PBAT-thermoplastic starch blends studied using solid-state NMR," in *International Journal of Polymer Analysis and Characterization*, vol 25, 2020, pp. 275-282.

[23] M. El H. El Nokab, et al, "Solid-state NMR spectroscopy: Towards structural insights into starch-based materials in the food industry," in *Polymers*, vol 14, 2022, pp. 4686.

[24] R. E. Dinnebier, S.J. L. Billinge, *Powder diffraction: Theory and Practice*, Cambridge, Royal Society of Chemistry, 2008, ch. 1.

[25] S. Ahmed, Annu, "Bionanocomposites in tissue engineering and regenerative medicine: Characterization of bionanocomposites," in *Woodhead Publishing Series in Biomaterials*, 2021, ch. 4.

[26] R. Narain, *Polymer science and nanotechnology: Fundamentals and applications: Nanomaterials characterization*, Elsevier Inc. 2020, ch. 17.

[27] J. D. Menczel, R. B. Prime, *Thermal analysis of polymers: Fundamentals and applications*, New Jersey, John Wiley & Sons, Inc., 2009, ch. 5.

[28] S. Patra, P. M. Ajayan, T. N. Narayanan, "Dynamic mechanical analysis in material science: The Novice's tale," in *Oxford Open Materials Science*, vol. 1, 2021, pp. 1.

[29] S. Thomas, R. Thomas, A. K. Zachariah, R. K. Mishra, *Spectroscopic methods for nanomaterials characterization: Nuclear magnetic resonance spectroscopy*, Elsevier Inc, 2017, ch. 17.

Human-Machine Collaboration in Industry 5.0

[1]*Maroš KRUPÁŠ (1st year),*
Supervisor: [2]Iveta ZOLOTOVÁ, Consultant: [3]Erik KAJÁTI

[1,2,3]Dept. of Cybernetics and Artificial Intelligence, FEEI TU of Košice, Slovak Republic

[1]maros.krupas@tuke.sk, [2]iveta.zolotova@tuke.sk, [3]erik.kajati@tuke.sk

***Abstract*—This paper describes the Industry 5.0 concept and its enabling technologies. With regard to this concept, the paper's main focus is also to provide an overview of the human-machine collaboration concept, which aims to enhance the working capabilities of humans and machines. In this case, new challenges for researchers and scientists arise when designing collaborative applications that put humans at the center of production while maintaining the flexibility and effectiveness of different applications. Our next research step is to design an appropriate framework and methods of AI and suitable enabling technologies to build these applications, which will tackle all the upcoming challenges.**

***Keywords*—Industry 5.0, Intelligent Systems, Human-Centric Manufacturing, Human-Machine Collaboration**

I. Introduction

The final form of the Industry 5.0 concept was introduced in 2021 [1] by the European Commission to shift the focus of industries to new values. Unlike its predecessors, this industrial revolution did not bring a big technological leap, and it is said to be a complement to the existing Industry 4.0 paradigm. This also results in a way that we should not take the fifth industrial revolution as the chronological continuation of the fourth, but rather as the co-existing one. Industry 5.0 is defined by its three key elements - human-centricity, sustainability, and resilience. These elements, also called core values of Industry 5.0, have brought different perspectives we should take when developing or implementing new technologies. It has also brought a need to develop already existing technologies further.

One of these areas that need to be further developed is the concept of human-machine collaboration, which focuses on collaborative solutions where human workers and machines share their skills and tasks. The new paradigm shift has brought a need to tackle all the challenges of using the best capabilities of robots and machines while keeping the focus on humans. Human-machine collaboration is becoming the new frontier in industrial applications and is set to be implemented in more applications in many areas in the future.

The first section describes the concept of Industry 5.0, its core values, and enabling technologies. The second section will focus on human-machine collaboration and analyze this concept and its applications in recent years. Subsequently, new approaches and challenges for human-machine collaboration are examined. The third section will be dedicated to future research directions.

II. Industry 5.0

The Industry 5.0 vision is to recognize the power of industry to achieve societal goals beyond jobs and growth, to become a resilient provider of prosperity by making production respect the boundaries of our planet and placing the wellbeing of the industry worker at the center of the production process [1]. The document also says that the main difference is that Industry 4.0 is technology-driven, while Industry 5.0 is value-driven. Industry 5.0 has its roots in Industry 4.0, which was also focused on societal goals and sustainability, but not as much on digitalization to increase the effectiveness and flexibility of manufacturing. The concept of Industry 5.0 has brought a different perspective, emphasizing the importance of research and innovation to support humanity in the long term. Industry 4.0 also addresses the issues of human-centricity, sustainability, and resiliency, but in the case of Industry 5.0, there is a shift to focus from individual technologies to a systematic approach [2].

While creating the concept of Industry 5.0, EU also identified its 6 key enabling technologies [3]. Each of the technologies introduced by the EU is believed to unfold its full potential only when combined with others to create complex systems. For that, we also need to develop a systematic approach [3]. These technologies are:

1) Individualised human-machine interaction - human-centric technologies that combine and interconnect the strengths of machines and humans. These technologies include multi-lingual speech and gesture recognition, human intention prediction, technologies to track the mental and physical state of the patients, collaborative robots, augmented, virtual, and extended reality, exoskeletons, or decision support systems.
2) Bio-inspired technologies and smart materials - They allow the creation of products with advanced capabilities from materials, which could be recycled, self-repaired, self-healed, or have embedded sensors technologies or biosensors.
3) Digital twins and simulation - technologies for modeling, controlling, and supervising the entire system to optimize production, test products, and processes, predict maintenance or failures, and measure environmental and social impact.
4) Data transmission, storage, and analysis technologies - cybersecurity, interoperability, edge computing, traceability, networked sensors, and big data management.
5) Artificial Intelligence - technologies capable of finding correlations and causality in dynamic systems, brain-machine interface, ability to respond to new situations without human support.
6) Technologies for energy efficiency, renewables, stor-

age, and autonomy - increasing energy demand of the mentioned technologies requires using renewable energy sources, low-energy data exchange, and smart dust to achieve emission neutrality.

These 6 enabling technological concepts also cross paths with nine enabling technologies of Industry 4.0, which were identified by BCG [4]. Therefore it is undoubtful that the technologies of Industry 4.0 will help to accomplish the goals of Industry 5.0. Furthermore, while these technologies were being discussed, there was a consensus that social and environmental needs must be better integrated into technology development.

The mentioned concepts have to support three key elements of Industry 5.0: human-centricity, sustainability, and resilience. Instead of taking new technology as a starting point and investigating its ability to increase efficiency, human-centric solutions place basic human needs and interests at the center of the production process. It also means ensuring that new technologies do not interfere with workers' fundamental rights, such as the right to privacy, autonomy, and human dignity. The industry should respect the limits of the planet and must be sustainable. It needs circular processes that reuse, re-evaluate and recycle natural resources, waste, and recycling environmental impact. Artificial intelligence and additive manufacturing technologies can also significantly optimize resource efficiency and minimize waste. Resilience means a higher level of resilience in industrial production to better arm it against external disruption and ensure that it can provide and support critical infrastructure in times of crisis. Geopolitical changes and natural crises such as the COVID-19 pandemic highlight the fragility of our current approach to globalized production.

III. Human-Machine Collaboration

Human-machine collaboration (HMC) is a term we use to describe humans and machines working together simultaneously in a shared workspace on the same tasks with possible physical contact. It involves parties with different capabilities, competencies, and resources, which must be coordinated to maximize their strengths. When we talk about human-machine collaboration in general, we do not limit the connotation of the word "machine", so in our case, it may refer to any physical equipment or software like an autonomous system, robot, algorithm, or AI. The first subsection of this chapter will focus on analyzing human-machine collaboration and its current research state. The second subsection analyzes the challenges of human-machine collaboration, while the third subsection focuses on my next work on this topic.

A. Review of HMC concepts and technologies

Since the introduction of Industry 4.0, there has been an increased focus on different types of human-machine interaction and collaboration. Most of the research on the Industry 4.0 focused on system-centric human-machine collaboration [5]. Some concepts created in the era of Industry 4.0 that already tried to shift their focus on a socially sustainable manufacturing workforce, for example, Operator 4.0 [6], in the context of human cyber-physical systems and human-automation symbiosis.

Before the introduction of Industry 5.0, many studies on human-robot collaboration focused on effective collaboration between humans and cobots. Subsequently, they led to the creation of human-robot collaborative assemblies [7], [8]. Cobots or collaborative robots are designed to interact physically with humans in a shared environment without barriers or protective cages. To collaborate with the cobot, we can use voice processing, gesture recognition, haptic interaction, brainwave reception, or deep learning for classification, recognition, and context awareness identification. These applications also proved useful to be extended and used with their digital twin (Fig. 1), which enabled real-time control and dynamic skill-based task allocation. Using predictions and simulations led to optimized behavior without the risk of human injury or financial loss [9], [10]. These human-robot collaboration applications were implemented and tested in manufacturing workspaces regarding their safety. However, these applications and studies were limited only to cobots, which are usually implemented in closed industrial cells.

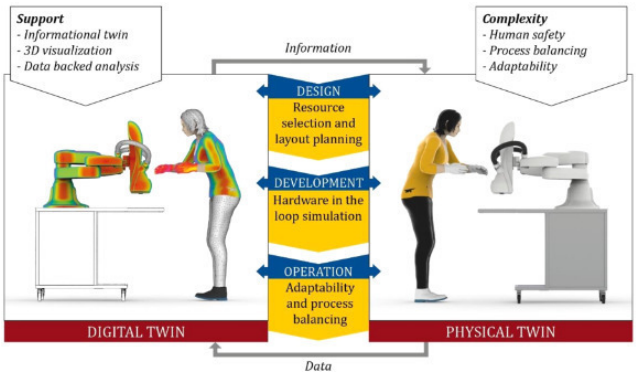


Fig. 1: Digital twin of human-robot collaboration [10]

New research regarding human-machine collaboration in the context of the Industry 5.0 focuses more deeply on its core values, making the collaboration human-centric, resilient, and sustainable. In this new industrial revolution, there has been a focus on new human-centric concepts of human-machine collaboration like human-machine symbiosis [11], [12], which focuses on mutual benefiting of humans and machines through the concept of emergence, where high-level, complex system emerges from the interaction of low-level system components.

Similarly, the industrial human needs pyramid (Fig. 2) was proposed [5], which classifies the human needs in the human-machine relationship in the industrial environment into five layers that follow the 5C journey - coexistence, cooperation, collaboration, compassion, and coevolution. Levels 1 to 3 of human needs have been primarily addressed by research in coexistence, cooperation, and collaboration relationships, while compassion and coevolution relationships, which need to be addressed in levels 4 and 5, could be the next direction of Industry 5.0 human-centric manufacturing. Based on this pyramid model, a human-centric human-machine collaboration framework in manufacturing was designed to extend standard human-robot collaborative practices to achieve human-centricity alongside production efficiency. This framework consists of three parts:

1) Dynamic human understanding - refers to developing a personalized digital twin of human wellbeing and modeling human states and intent.
2) Empathetic robot control - proactive human assistance, which involves the generation of empathetic action based on changing human behavior or working preferences.

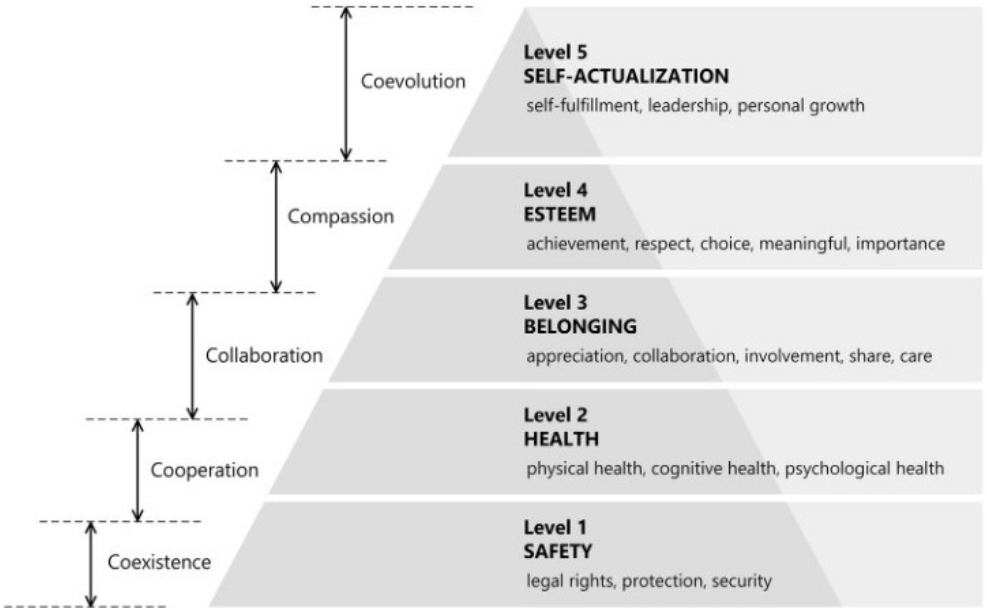


Fig. 2: Human-machine relationship levels mapped to Industrial Human Needs Pyramid [5].

3) Dynamic task scheduling - dynamic human-robot task scheduler formed through collaborative intelligence.

Researchers found that companies achieve the most significant performance when humans and AI join forces together to create collaborative intelligence [13]. Through collaborative intelligence, humans and AI can actively enhance each other strengths. Humans can are better at leadership, creativity, and social skills, while AI could improve speed, scalability, and quantitative capabilities.

When designing human-machine collaboration in Industry 5.0 concept, there is also a focus on resilient human-machine teams and collaboration within them. The human-machine team is a team that consists of human actors with different skills, collaborative robots, and AI-based systems for supervision or assistance. One of the main concerns of resiliency can include task allocation to utilize the best capabilities of all actors in human-machine teams. These teams of humans and machines need to adapt to changing situations where the division of work and responsibility must be altered smoothly [14].

B. Challenges of Human-machine collaboration

Designing human-machine collaboration according to the new principles of the Industry 5.0 concept has brought new opportunities and challenges. Before these challenges are solved, they need to be identified and described first. Based on the reviewed literature [15], [16], [17], [18], [3], [5], we put some of these identified technical and social challenges into Table I. Overcoming these challenges would result in improved human well-being and flexible manufacturing, where humans and machines develop their capabilities [11].

In [16], authors proposed a framework with guidelines and recommendations for three complexity levels of the influencing factors presented when designing human-centered human-machine interaction workspaces in an industrial setting. It was concluded that challenges from designing human-machine collaboration workspaces in industrial settings require multidisciplinary and diverse knowledge of fields with a framework to systematize research findings.

Authors in [18] addressed decision-making risks while using AI or other algorithms in human-machine collaboration. Humans are vulnerable to cognitive biases, while machines cannot handle incomplete information well. In their paper, the authors proposed research challenges covering organizing efficient and healthy human-machine teams with effective mental interfaces.

Having humans working with cobots seems an efficient way to develop personalized products, but it creates issues and considerations we should address. As described in [17], these issues may include human fear of losing jobs, psychological concerns, or dynamic task allocation. The authors described the human-robot co-working in the factory as better when cobots complement humans in repeated activities and let humans do creative and innovative jobs.

C. Research gap

Based on our research, we believe there is a need to design a new approach for developing human-machine collaboration applications, which would also be suited to use with heterogeneous multi-robotic systems. Our focus is that it would also apply to different mobile robots, including unnamed ground and aerial vehicles working simultaneously in the same shared workspace. The expected result of my work is to design a conceptual framework for different types of human-centered solutions which focuses on human-machine collaboration applications. This framework will be subsequently tested in laboratory conditions using physical devices. Our goal is to also focus on implementing the Industry 5.0 enabling technologies while focusing on Industry 5.0 core values and addressing the challenges of designing human-centric applications. We would use a diploma thesis application focused on laboratory multi-robotic systems as a starting point.

My diploma thesis mainly focused on machine-machine and human-machine cooperation between aerial and ground unmanned vehicles. The terms cooperation and collaboration are sometimes used interchangeably in human-machine interaction research, but they are different types of interactions [19]. In human-machine cooperation, humans and machines

TABLE I: Challenges of human-machine collaboration in Industry 5.0

HMC Challenge	Description
Human-centric AI	AI personalized to individual workers, focusing on enhancing human productivity rather than replacing humans.
Transparency, Explainability	Understanding of the machine and AI decisions and reasons behind their actions.
Technology acceptance and trust	Human fear of losing jobs and psychological concerns.
Training People	Continuous up-skilling or re-skilling to ensure smooth human-machine interaction performance.
Safety	Safe collaboration and anticipation of potential emergencies
Performance measures	Metrics and means of measuring human and machine wellness to define the quality of collaboration.

work on common goals but do not share tasks. This is common in manufacturing, where robots start to do one part of the manufacturing process, and humans finish the process afterward. My diploma thesis focused on creating an application to control cooperation between human, drone, and ground mobile robot. The result of the thesis was a mobile application that human operators can use to control the devices. It gives them tasks via an interface with buttons and a map on the screen, including ground robot navigation or drone landing on this ground mobile robot.

Machines in real industry applications should be capable of responding to unexpected or new situations while making real-time decisions when working with human operators and other machines. Most of the human-machine collaboration frameworks proposed by other authors focus on what we need to focus on regarding human-centricity and resiliency, but not on the specific use of all Industry 5.0 enabling technologies in one framework to create a complex system. For this process, we need to design a framework capable of gathering, processing, and analyzing a large amount of data. Both humans and machines will generate a large amount of data, which will be gathered and processed. We will use this data to make simulations and predictions using the digital twin of the workspace and human-machine teams. Some decisions must be processed or decided on the cloud; others must be processed on edge devices to accomplish real-time responses to different situations. There will be a need to develop personalized and human-centric AI.

Our identified research gaps we would like to focus on in the next steps are, therefore:

1) Identify and implement technologies and methods suitable for the different types of human-centered solutions. This could include improving enabling technologies to their full potential when designing personalized AI or digital twin of HMC.
2) Design dynamic task allocation plan between humans and machines in HMC. This would result in better flexibility and resilience of the human-machine team.
3) Design a systematic approach to building safe and trustable robots using appropriate technologies and materials.

IV. Future Work

Many technologies and concepts mentioned in this paper could be used when developing future human-machine collaboration applications. In our next research, we would like to analyze methods and approaches of AI further and enable IT technology suitable for different types of human-machine collaboration applications that will follow Industry 5.0 principles. As part of our work, we could further explore augmented reality when interacting with machines, create a digital twin of both human and machine, implement human-centric and transparent AI, and conclude the results using appropriate metrics.

Acknowledgment

This publication was supported by the grant AI Based Approaches and Enabling IT Technology for Human-Machine Collaboration in Industry 5.0 (07/TUKE/2022) and VEGA 1/0480/22, EDEN: EDge-Enabled intelligeNt systems.

References

[1] E. C. D.-G. for Research and Innovation, "Industry 5.0 - towards a sustainable, human-centric and resilient european industry," 2021.
[2] X. Xu, Y. Lu, B. Vogel-Heuser, and L. Wang, "Industry 4.0 and industry 5.0—inception, conception and perception," *Journal of Manufacturing Systems*, vol. 61, pp. 530–535, 2021.
[3] E. C. D.-G. for Research and Innovation, "Enabling technologies for industry 5.0," 2020.
[4] B. C. Group, "Industry 4.0," 2023. [Online]. Available: https://www.bcg.com/capabilities/manufacturing/industry-4.0
[5] Y. Lu, H. Zheng, S. Chand, W. Xia, Z. Liu, X. Xu, L. Wang, Z. Qin, and J. Bao, "Outlook on human-centric manufacturing towards industry 5.0," *Journal of Manufacturing Systems*, vol. 62, pp. 612–627, 2022.
[6] D. Romero, J. Stahre, and M. Taisch, "The operator 4.0: Towards socially sustainable factories of the future," p. 106128, 2020.
[7] L. Wang, R. Gao, J. Váncza, J. Krüger, X. V. Wang, S. Makris, and G. Chryssolouris, "Symbiotic human-robot collaborative assembly," *CIRP annals*, vol. 68, no. 2, pp. 701–726, 2019.
[8] E. Magrini, F. Ferraguti, A. J. Ronga, F. Pini, A. De Luca, and F. Leali, "Human-robot coexistence and interaction in open industrial cells," *Robotics and Computer-Integrated Manufacturing*, vol. 61, p. 101846, 2020.
[9] A. Bilberg and A. A. Malik, "Digital twin driven human–robot collaborative assembly," *CIRP annals*, vol. 68, no. 1, pp. 499–502, 2019.
[10] A. A. Malik and A. Brem, "Digital twins for collaborative robots: A case study in human-robot interaction," *Robotics and Computer-Integrated Manufacturing*, vol. 68, p. 102092, 2021.
[11] C. S. e. a. Lu Y., Adrados J., "Humans are not machines—anthropocentric human–machine symbiosis for ultra-flexible smart manufacturing," *Engineering*, vol. 7, pp. 734–737, 2021.
[12] Y. Yao, "Human-machine co-intelligence through symbiosis in the smv space," *Applied Intelligence*, pp. 1–21, 2022.
[13] H. J. Wilson and P. R. Daugherty, "Collaborative intelligence: Humans and ai are joining forces," *Harvard Business Review*, vol. 96, no. 4, pp. 114–123, 2018.
[14] E. Kaasinen, A.-H. Anttila, P. Heikkilä, J. Laarni, H. Koskinen, and A. Väätänen, "Smooth and resilient human–machine teamwork as an industry 5.0 design challenge," *Sustainability*, vol. 14, no. 5, p. 2773, 2022.
[15] S. Nahavandi, "Industry 5.0—a human-centric solution," *Sustainability*, vol. 11, no. 16, p. 4371, 2019.
[16] S. J. e. a. Simões A., Pinto A., "Designing human-robot collaboration (hrc) workspaces in industrial settings: A systematic literature review," *Journal of Manufacturing Systems*, vol. 62, pp. 28–43, 2022.
[17] B. P. e. a. Maddikunta P., Pham Q., "Industry 5.0: A survey on enabling technologies and potential applications," *Journal of Industrial Information Integration*, vol. 26, 2022.
[18] M. L. e. a. Xiong W., Fan H., "Challenges of human—machine collaboration in risky decision-making," *Frontiers of Engineering Management*, vol. 9, pp. 89–103, 2022.
[19] A. Kolbeinsson, E. Lagerstedt, and J. Lindblom, "Foundation for a classification of collaboration levels for human-robot cooperation in manufacturing," *Production & Manufacturing Research*, vol. 7, no. 1, pp. 448–471, 2019.

Measuring and Monitoring Particulate Matter in an Industrial Environment

[1]*Simona KIREŠOVÁ (2nd year)*
Supervisor: [2]Milan GUZAN

[1,2]Dept. of Theoretical and Industrial Electrical Engineering, FEI TU of Košice, Slovak Republic

[1]simona.kiresova@tuke.sk, [2]milan.guzan@tuke.sk

***Abstract*— Particulate matter are small particles that have negative effects on human health. In this paper, we outline the measurement and monitoring system for particulate matter used in an industrial environment – a small company that is focused on producing medical supplies, which must meet the conditions of ISO Class 8 of air cleanliness as defined by the STN EN ISO 14644-1 standard. We point out the benefits of long-term monitoring as opposed to once-per-year measurement.**

***Keywords*—Classes of air cleanliness, ESP8266, InfluxDB, number concentration, particulate matter, SPS30**

I. Introduction

Particulate matter (PM) are small particles of varying size: PM10 – particles with a diameter <10 μm, PM2.5-10 or coarse particles (2.5–10 μm), PM2.5 or fine particles (<2.5 μm), and PM0.5 or ultrafine particles (<0.5 μm) [1]. PM can be described by their mass concentration, number concentration, surface area, and size distribution. Mass concentration measurements are generally used for the estimation of the health effects of PM. Number concentration and size distribution measurements are usually used to assess the impact of PM on the climate [2], but they are also useful in measuring ultrafine particles [3].

Measuring PM is important due to the negative effects of the particles on human health on the cardiovascular and respiratory systems. Exposure to PM can cause decreased lung function, difficulty breathing, asthma, high blood pressure, irregular heartbeat, and heart attack [4]-[6]. The health effects of PM are more pronounced with decreasing size since fine particles can penetrate deeper into the respiratory system and reach lung alveoli and ultrafine particles can reach the bloodstream [3].

In our previous research, we measured the concentrations of particulate matter in urban and rural areas from short-term [7] and long-term [8] point of view. We have also analyzed how the distance from the source of PM affects the mass concentration [9], as well as how often the measurements should be carried out [10]. In [11] we focused on determining the correlation between PM mass concentration and meteorological factors such as temperature, humidity, and atmospheric pressure. In [12] we focused on measuring fine and ultrafine particles. However, all those measurements were focused on outdoor particulate matter concentrations. Papers [13] and [14] serve as a more detailed overview of methods of measurement of PM. However, all those measurements were focused on measuring PM outdoors.

This paper will be focused on measuring PM indoors, in a controlled environment in a company, which must meet certain conditions of air quality.

II. Measuring PM in an Industrial Environment

A. Classes of Air Cleanliness

The classes of air cleanliness are defined in the STN EN ISO 14644-1 standard [15]. There are 9 classes of air cleanliness by particle concentration defined by the standard. Each class is defined by the maximum allowable number concentration of particles (in #/m^3) for particles of various sizes in Table 1. The stricter classes are defined by smaller particles, as the sampling and statistical limitations make the measurement of larger particles inappropriate. On the other hand, the less strict classes are defined by the number concentration of larger particles, as the concentration limits of smaller particles are not applicable due to high particle number concentration.

TABLE I
ISO Classes of Air Cleanliness by Particle Size [15].

ISO Class	Max. concentration (#/m^3) for particles of equal or greater size:					
	0.1 μm	0.2 μm	0.3 μm	0.5 μm	1 μm	5 μm
1	10	-	-	-	-	-
2	100	24	10	-	-	-
3	1000	237	102	35	-	-
4	10000	2370	1020	352	83	-
5	100000	23700	10200	3520	832	-
6	10000000	237000	102000	35200	8320	293
7	-	-	-	352000	83200	2930
8	-	-	-	3520000	832000	29300
9	-	-	-	35200000	8320000	293000

The measuring of PM took place in a small company in Slovakia that focuses on the production of medical tools such as gauze, bandages, disposable sterile surgical operation sheets, and medical masks. This company must meet the conditions of ISO Class 8 (highlighted by grey color in Table 1) in the production hall, where 21 HEPA filters are installed to ensure the required air cleanliness. This means that the limit value for PM ≥ 0.5 μm is 3 520 000 particles per cubic meter (#/m^3). Once per year, the company undergoes a certification when PM is measured. The floor plan of the production hall is depicted in Fig. 1, where the positions of the HEPA filters are marked as F1—F21 and two measurement devices for particulate matter as M1 and M2 (they will be described in more detail in Section II.B).

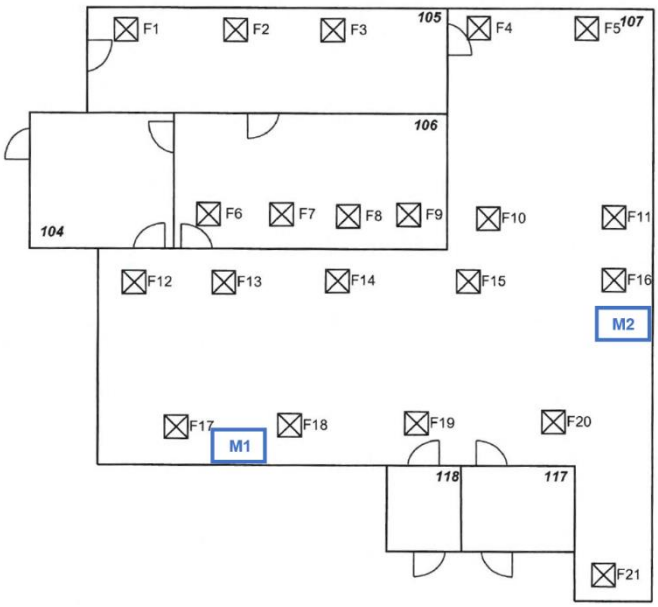


Fig. 1. The floor plan of the production hall with the marked positions of HEPA filters and PM measurement devices.

B. The Measurement System

The measurement system was based on the Wemos D1 R2 development board with the ESP8266 microcontroller. The block diagram of the measurement system is shown in Fig. 2.

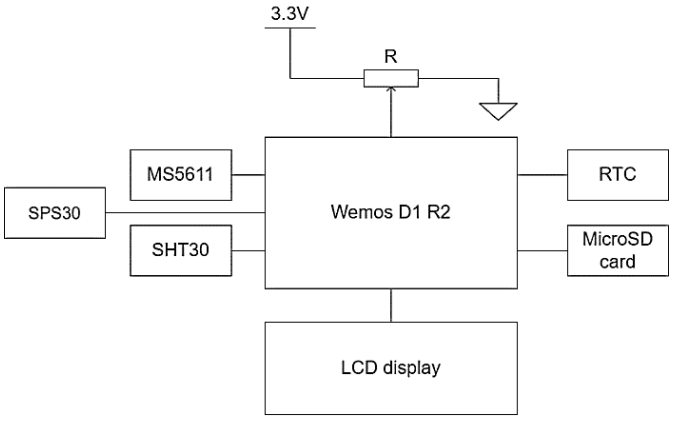


Fig. 2. The block diagram of the measurement system.

The sensors used in the systems were the PM sensor SPS30, the temperature and humidity sensor SHT30, and temperature and atmospheric pressure system. SPS30 is an optical particle sensor. It is based on the principle of laser diffraction. SPS30 detects the intensity and angle of a laser beam after it has been struck and scattered by a particle. An algorithm then determines mass concentration, number concentration, and the typical size of the particles. It measures PM0.5, PM1, and PM2.5 with an accuracy of ±10% and PM4 and PM10 with an accuracy of ±25% [16].

Sensor SPS30 was chosen because of its large range of measured particle sizes, availability on the Slovak market, its cost, as well as the possibility to interface the sensor with the Wemos D1 R2 development board. At the time of the design and implementation of the measurement system, SPS30 was one of the few low-cost PM sensors available on the Slovak marker. Of course, since then, as more PM sensors found their way to the market, have also been considered. For example, sensors SEN5x (SEN50, SEN54, SEN55) can measure mass concentration of PM1, PM2.5, PM4 and PM10 as well as number concentration of PM0.5, PM1, PM2.5, PM4 and PM10, however the accuracy of SEN5x sensors is the same as SPS30 (±10% for PM0.5, PM1 and PM2.5 and ±25% for PM4 and PM10). However, some of the SEN5x sensors can also measure other physical quantities, such as temperature, humidity, VOC index (SEN54 and SEN55), and NOx index (SEN55) [17]. Sensor HPMA115C0-003 can measure PM2.5 with an accuracy of ±15% and PM1, PM4, and PM10 with an accuracy of ±25% [18], which is worse than the accuracy of SPS30. Other PM sensors do not provide the measurements of multiple PM size categories. For example, sensor HPMA115S0-XXX can only measure PM2.5 and PM10 with an accuracy of ±15% [19]. Several months after the deployment of the measuring system, a new type of PM sensor arrived on the Slovak market, sensor IPS-7100. It can measure more categories of PM such as PM0.1, PM0.3, PM0.5, PM1, PM2.5, PM5, and PM10, all with an accuracy of ±10%, thus improving the measurement accuracy of the coarse particles [20]. This sensor was later implemented into a measurement system for outdoor monitoring of PM and meteorological factors, as it was described in [12].

We use sensors SEN54, SEN55, and IPS-7100 in other measurement systems, deployed for outdoor measurements in urban and rural areas.

As for the rest of the measurement system, the measured data is logged in a *.csv file on the microSD card every 5 seconds. Real-Time Clock module timestamps new entries with the date and time of the measurement.

All the measured values of PM concentrations, temperature, humidity, and pressure are periodically sent to the InfluxDB database via WiFi. A potentiometer is included in the measurement system for adjusting the interval of sending data to the InfluxDB cloud. This interval ranges from 30 seconds to 10 minutes and it is independent of the 5-second measurements recorded on the microSD. This interval of sending the measurements to the database is currently set by the potentiometer as once per minute and it is less frequent than the 5-second SD card logging interval to reduce the amount of data sent to the InfluxDB database, while still retaining enough measurements to track the dynamic changes in PM concentration in real-time. However, we use the data logged on the microSD card for a more thorough analysis. Several dashboards were created in the database, which display the graphs of the measured physical quantities.

Some of the measured values (not all) are displayed on the 20x4 LCD display, namely, the number concentration of PM with diameter ≥ 0.5 µm, temperature, humidity, and pressure, along with the current date and time and the interval of sending data to InfluxDB set by the potentiometer (this was set for 60-seconds after the prototype of the measurement system was assembled).

The case was modeled in Fusion 360. It comes in two parts – the case and the cover, which are put together using screws. The case includes openings for the sensor on one side and for the SD card as well as USB and power cables on the other side. It also contains vent holes in four of its sides. The cover contains an opening for the display, a potentiometer, and screws. Both parts were 3D printed using PLA. Two prototypes of the measurement system were assembled and one of them is shown in Fig. 3.

Fig. 3. The prototype of the measurement system.

C. The Results

The results are obtained from the measurements which were taken every 5 seconds and logged on the microSD card in *.csv files. They were processed in MATLAB, where the hourly averages were calculated, and the graphs were plotted. These measurements are long-term. They started on August 10, 2022, and they are still ongoing. To confirm the effectiveness of the filtration system from a more long-term point of view, hourly averages of PM0.5-10 number concentration from 10/8/2022 to 14/10/2022 were plotted in the same graph in Fig. 4. Different colored lines represent measurements during each day. The x-axis on the graph represents hours in a day (date not specified, since we want to compare the daily measurements). The horizontal red dashed line represents the limit value set by STN EN ISO 14644-1. Vertical black dashed lines mark 6:00 and 14:00 (the start and end of the working time in the company). Only the workdays were plotted in Fig. 4 (measurements during the weekends and holidays were ignored). After the weekends and holidays were filtered out, a total of 45 days of measurements during workdays remain in Fig. 4, from which it can be seen that the number concentration decreases significantly for the duration of working hours (6:00 – 14:00, as indicated by the black dashed lines) when the filtration system is active. In all 45 measurements, the number concentration of PM fell under the limit value (red dashed line).

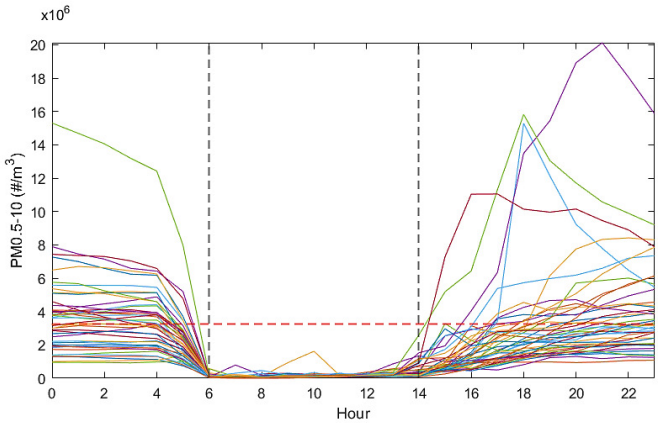


Fig. 4. Hourly averages of PM0.5-PM10 number concentration measured during the workdays from 10/8/2022 to 14/10/2022 (a total of 45 measured days represented by different colored lines).

Similarly, Fig. 5 shows the hourly averages of typical particle size for the same 45 workday measurements (like in Fig. 4) represented by the different colored lines, which are plotted against hours in the day on the x-axis. Two black dashed lines mark 6:00 and 14:00. It is apparent from Fig. 5, that while the filtration system is off (before 6:00), the typical particle size tends to be between 0.4 and 0.5 µm. As soon as the filtration system is turned on at 6:00, typical particle size starts increasing (usually right until it is turned off). There is more variation between typical particle sizes between 6:00 and 14:00 – the particle size ranges from over 0.5 µm to over 1.1 µm. However, as the filtration is turned off at 14:00, we can observe a sharp decrease in typical particle sizes back in the ranges similar to their morning/pre-filtration levels.

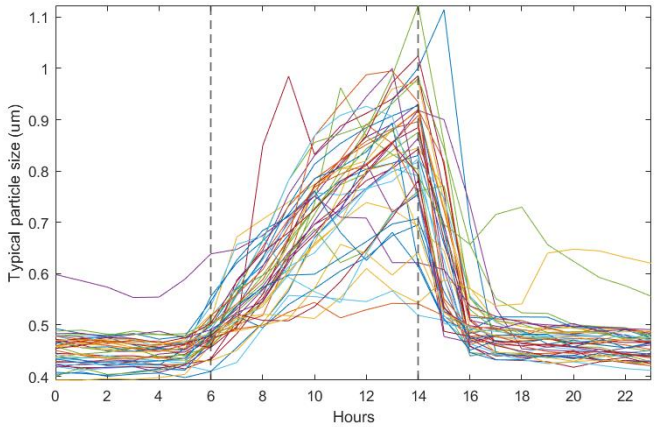


Fig. 5. Hourly averages of typical particle size measured during the workdays from 10/8/2022 to 14/10/2022 (a total of 45 measured days represented by different colored lines).

III. Conclusion

For the vast majority of the working time, the conditions in the company regarding particulate matter meet the Class 8 standard set by STN EN ISO 14644-1. The effect of the filtration system is apparent, in all 45 days of measurement displayed in Fig. 4 and Fig. 5. Still, long-term monitoring of PM is invaluable for keeping the necessary conditions in controlled spaces. Not only does monitoring of PM provide an overview of the current working conditions, but it can also help alert us about failures in the filtration system faster. Long-term monitoring can also alert the company for a possible subtler, long-term deterioration in the functionality of the filtration system, which might not be uncovered as easily otherwise. More measuring devices need to be designed and realized. Two storage rooms and one smaller production hall also need to be monitored. With additional devices, all bases of production and storage will be fully monitored. Furthermore, while InfluxDB offers visualizations of the measured data, an application for the processing, analysis, and visualization of the data logged in the *.csv file would also be valuable, as the *.csv file contains data measured every 5 seconds (as opposed to data in InfluxDB what was sent to the database in 60-second intervals).

This part of my dissertation focused only on measuring and monitoring PM in an industrial environment. Another part of my dissertation is focused on PM measurement in a home environment and outdoor measurements in urban and rural areas. Of course, different places of measurement will have different conditions that will affect the levels of PM concentration and air quality. In outdoor measurements, meteorological factors such as temperature, humidity, atmospheric pressure, and wind speed become important factors that might influence the concentrations of PM. Human activity such as wood combustion, construction, and traffic also impacts air quality in residential areas. If a network of

measurement stations situated in different locations is created, we will be able to see how PM correlates with meteorological factors. Some (but not all) of the outdoor measurement stations in this network have been deployed. Based on the measured data, both linear and non-linear correlation of several variables need to be investigated since the relationship between the variables may not be linear. Further analyses are also possible using cloud computing in InfluxDB, e.g., calculating hourly or daily averages, or calculating the Air Quality Index (AQI), which is a parameter with a more informative value about the air quality. Machine learning is another tool that could be used for creating a prediction model of PM based on the other measured physical quantities.

References

[1] R. M. Harrison et al.: "Airborne particulate matter: sources, atmospheric processes, and health," Royal Society of Chemistry, UK, 2016, ISBN: 978-1-78262-491-2.

[2] S. D. Lowther et al: "Particulate Matter Measurement Indoors: A Review of Metrics, Sensors, Needs, and Applications," Environmental Science & Technology, vol. 53, no. 20, pp. 11644-11656, Oct 2019.

[3] H. Kwoon, M. Ryu, C. Carlsten: "Ultrafine particles: unique physiochemical properties relevant to health and disease," Experimental and Molecular Medicine, vol. 52, pp. 318-328, Mar 2020.

[4] A. Suleiman, M. R. Tight, A. D. Quinn: "Assessment and prediction of the impact of road transport on ambient concentrations of particulate matter PM10," Transportation Research Part D: Transport and Environment, vol. 49, pp. 301-312, Dec. 2016.

[5] C. Mühlfeld et al.: "Interactions of nanoparticles with pulmonary structures and cellular responses," The American Journal of Physiology-Lung Cellular and Molecular Physiology, vol. 294. no. 5, pp. L817-L829, May 2008.

[6] M. Kampa, E. Castanas: "Human health effects of air pollution," Environmental Pollution, vol. 151 no. 2, pp. 362-367, Jan 2008.

[7] S. Kirešová, M. Guzan, P. Galajda: "Measuring Particulate Matter (PM) Using SPS30," In 32nd International Conference Radioelektronika, pp 160–165. Košice, Slovakia, Apr 2022.

[8] S. Kirešová, M. Guzan, V. Rusyn: "Particulate Matter PM2.5 and PM10 and Its Impact on Air Quality in Urban or Rural Areas," In 2nd International Workshop on Information Technologies: Theoretical and Applied Problems. Ternopil, Ukraine, Nov 2022.

[9] S. Kirešová, M. Guzan: "Concentration of Particulate Matter in the Air: A Comparison of Measurements," Electrical Engineering and Informatics 13, pp. 301-306, Jul 2022. (In Slovak).

[10] S. Kirešová, M. Guzan: "Analyzing the Sample Period of Particle Mass Concentration Measurements," In Proceeding of scientific and student's works in the field of Industrial Electrical Engineering 11, pp. 210-214. Košice, Slovakia, Jul 2022.

[11] S. Kirešová, M. Guzan: "Determining the Correlation between Particulate Matter PM10 and Meteorological Factors," Eng, vol. 3, no. 3, pp. 343-363, Aug 2022.

[12] S. Kirešová, M. Guzan, B. Sobota: "Using Low-Cost Sensors for Measuring and Monitoring Particulate Matter with a Focus on Fine and Ultrafine Particles," Atmosphere, vol. 14, no. 2, p. 324, Feb. 2023.

[13] S. Kirešová: "Particulate Matter and the Methods of Its Measurement: An Overview," In SCYR 2022: 22nd Scientific Conference of Young Researchers, Proceedings from Conference, pp 43-47, Apr 2022.

[14] S. Kirešová, M. Guzan: "Measurement of Particulate Matter: Principles and Options of Measurement at Present," Acta Electrotechnica et Informatica, vol. 22, no. 2, pp. 8-18, Jun 2022.

[15] "Cleanrooms and associated controlled environments - Part 1: Classification of air cleanliness by particle concentration." ISO Standard no. 14644-1, International Organization for Standardization, 2015. [Online]. Available: <https://www.iso.org/standard/53394.html>. (Accessed 2022-10-12).

[16] Sensirion. "Particulate Matter Sensor for Air Quality Monitoring and Control," Datasheet SPS30. Mar 2020. [Online]. Available: <https://cdn.sos.sk/productdata/98/89/92718144/sps30-2.pdf>. (Accessed 2022-05-30)

[17] Sensirion. "Environmental Sensor Node for HVAC and Air Quality Applications," Datasheet SEN5x. Jan 2022. [Online]. Available: <https://cdn.sparkfun.com/assets/5/b/f/2/8/Sensirion_Datasheet_SEN5x.pdf>. (Accessed 2022-05-30).

[18] Honeywell. "Particulate Matter Sensors, Issue F," Datasheet HPM Series. May 2019. [Online]. Available: <https://www.farnell.com/datasheets/2829051.pdf>. (Accessed 2023-04-04).

[19] Honeywell. "Particulate Matter Sensors, Issue A," Datasheet HPM Series. Nov 2016. [Online]. Available: <https://www.farnell.com/datasheets/2313714.pdf>. (Accessed 2023-04-04).

[20] Piera Systems. "Photon Counting Intelligent Particle Sensor for Accurate Air Quality Monitoring," Datasheet IPS Series Sensor. 2021. [Online]. Available: <https://www.farnell.com/datasheets/3749927.pdf>. (Accessed 2023-04-04).

The new routing strategy using artificial intelligence methods in Cloud MANET

[1]*Natalia KURKINA (2nd year)*
Supervisor: [2]Ján PAPAJ

[1,2]Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

[1]natalia.kurkina@tuke.sk, [2]jan.papaj@tuke.sk

***Abstract*—The development of mobile ad hoc networks and the popularization of cloud technologies led to the emergence of Cloud MANET. The new approach for routing-based artificial intelligence in Cloud MANET is proposed in this paper. The experimental part of the article compares the standard AODV reactive routing protocol and the AODV protocol with an embedded cloud. According to the comparison results, implementing cloud technologies increases the packet delivery ratios and reduces packet loss and end-to-end delay.**

***Keywords*— MANET, cloud technologies, Cloud MANET, routing protocols, AODV.**

I. Introduction

MANET (mobile ad hoc network) is decentralized, without infrastructure and has a constantly changing topology due to the high mobility of nodes. Each node in the network acts as a router and can forward information to other nodes. The use of such networks can be useful in areas where fixed infrastructure does not exist or has been destroyed, for example in areas affected by earthquakes [1].

Cloud MANET combines two areas - mobile ad hoc networks and cloud technologies. The main goal of such a combination was to achieve a high data transfer rate and reduce latency [2]. In Cloud MANET, in addition to mobile nodes in the network, the cloud also participates in decision-making and traffic redirection [3]. All mobile nodes in the network are grouped into clusters. Each node can belong to only one cluster at a time. One node is selected as the cluster head (CH) in each cluster). It is usually responsible for connecting to the cloud. To transmit information from one cluster to another, the information is first sent to the cloud, then the cloud sends the data to the corresponding CH of the other cluster, which forwards it to the destination node [4].

Link quality and mobility are major routing issues in MANET due to frequent disconnections and the need to update routes in real time. Therefore, the second main task is the selection of the optimal routing algorithm.

II. Proposed Method of Routing in Cloud MANET

This section describes the proposed design of the routing model in Cloud MANET.

When we talk about Cloud MANET, then there can be a cloud in the network that receives all information about nodes and paths. The network can also contain several MANET networks, each connected to a personal cloud, and these clouds will already be able to communicate with each other. In this case, nodes from one MANET network will be able to communicate with nodes from another MANET network through clouds. In the absence of the cloud, the network operates in the same mode based on standard routing using artificial intelligence methods. The basic idea of the new routing algorithm is shown in Fig. 1. So artificial intelligence methods will be helped for routing in MANET without present the Cloud, within the Cloud and in case of two Cloud communications.

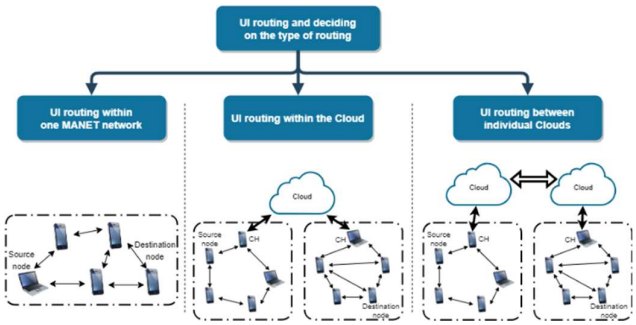


Fig. 1. The basic idea of the routing algorithm in Cloud MANET.

Thus, when the cloud is presented, it will know about all the nodes that are within reach of any other node connected to the cloud directly or through another node.

The principle of operation of the proposed algorithm is described below and in pseudocode 1.

PSEUDOCODE I
THE PRINCIPLE OF OPERATION

Algorithm 1. The principle of operation of the proposed algorithm

Input: Set of nodes

1. Start routing:
2. **If** (the node has a route to the destination in the routing table):
   Then Sends the packet to the destination.
3. **Else if** (the node has a connection to the cloud) **and** (the node is a CH):
   The node sends a packet to the cloud and Cloud chooses the appropriate path using the AI.
4. **Else if** (the node has a connection to the cloud) **and** (the node is not a CH):
   The node sends a packet to the CH and Cloud chooses the appropriate path using the AI.
5. **Else:**
   The node sends the packet using a standard routing protocol.

Output: Routing path to the destination

1. When a node wants to send a packet to the destination, it must first find out whether the destination is its neighbor and

does not already know the route to the destination. If so, it sends directly to the target node.

2. If he does not know the way to the destination, he must verify that he is not connected to the cloud. If it is, it will send it directly to the cloud, which will decide either to send the packet to another CH, or another cloud, and it will decide on the next path of the packet.

3. If it is not connected to the cloud, it must choose a CH using artificial intelligence. If there are no nodes among the known nodes that would be connected to the cloud, it sends the packet using the standard routing protocol.

4. If the node chooses CH, it sends a packet to CH. If the CH does not know about the route to the destination, it sends the packet to the cloud and thus point 2 of the algorithm is repeated.

Thus, the use of cloud technologies in combination with decentralized MANET can reduce route search time, and network latency and lead to increased throughput. The implementation of artificial intelligence methods in the CH selection and routing process will be the basis for further research.

III. Experimental Study

NS-3 simulator was used for Cloud MANET simulation [5]. In this paper, a cloud environment was added to reactive routing protocols AODV. Table I shows the simulation parameters. There are 2 scenarios simulation, one of them is the network which consists AODV routing protocols between MANET nodes and the cloud. The second scenario is the network which consists of only AODV routing protocols between MANET nodes. In both scenarios, some nodes are selected that exchange packets with each other.

TABLE I. Simulation Parameters

Parameter	Value
Simulation Area	300 x 300 m^2
Simulation Time	20 s
Number of MANET nodes	50
Max. transmission range of MANET node	50 m
Channel type	Wireless
Mobility Model	Random waypoint
Moving Speed	20 m/s

Performance metrics

The following performance metrics were selected to evaluate the proposed simulation scenario: packet delivery ratio, packet loss ratio and end-to-end delay [6].

Comparative study

Fig. 2 shows a comparison of packet delivery ratios of the two scenarios (AODV and AODVCloud). From the comparison, we conclude that scenarios AODV with cloud achieve a better packet delivery ratio than scenarios with only AODV routing protocol.

Also Fig. 2 shows a comparison of packet loss ratios of the two scenarios. From the comparison, AODV with cloud achieves a lower packet loss ratio. It is 20% better than in scenarios with only AODV routing protocol.

The comparison of the end-to-end delays and throughput of the two scenarios (AODV and AODVCloud) is showed in Fig. 3. From the comparison, we conclude that scenarios AODV with cloud achieve a lower end-to-end delay and higher throughput than scenarios with only AODV routing.

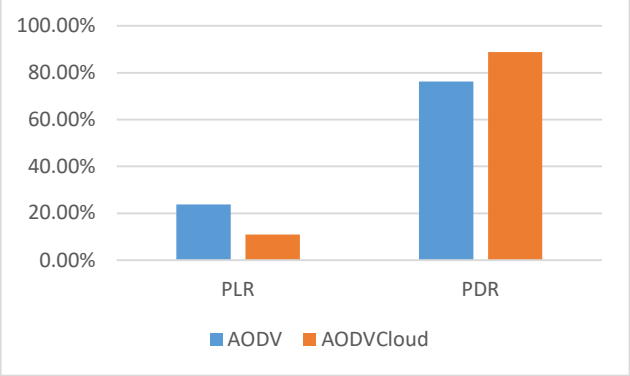


Fig. 2. Comparison of packet delivery and packet loss ratios

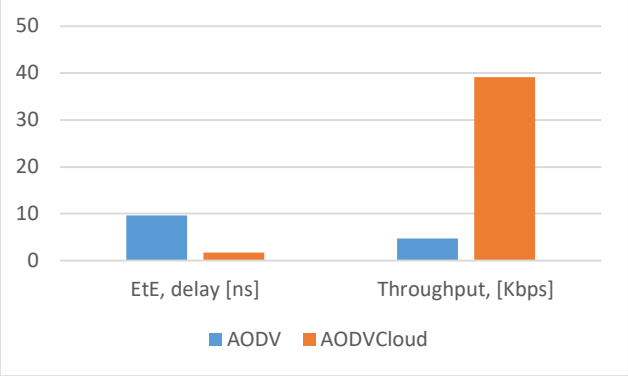


Fig. 3. Comparison of end-to-end delay and throughput

IV. Conclusion

In this article, we proposed a new model in Cloud MANET, which will be implemented in subsequent works. In conclusion, a comparative analysis of two network operation scenarios was presented - with and without an embedded cloud in the AODV. In future work, we plan to implement the presented routing scheme, and artificial intelligence methods in the CH selection and routing process and compare it with existing models.

Acknowledgement

This research was funded by the Slovak Research and Development Agency, research grant no. APVV-17-0208 and VEGA 1/0260/23.

References

[1] K. S. Sankaran, N. Vasudevan, K. R. Devabalaji, et al., "A Recurrent Reward Based Learning Technique for Secure Neighbor Selection in Mobile AD-HOC Networks," in IEEE Access, vol. 9, 2021.

[2] S. K. Singh and J. Prakash, "Energy efficiency and load balancing in manet: A survey," in 2020 6th International Conference on Advanced Computing and Communication Systems, pp. 832–837, 2020.

[3] S.A. Alghamdi, "Novel trust-aware intrusion detection and prevention system for 5g manet–cloud," International Journal of Information Security, 2021.

[4] S.A. Alghamdi, "Three-Tier Architecture Supporting QoS Multimedia Routing in Cloud-Assisted MANET with 5G Communication (TCM5G).", Mobile Networks and Applications, vol. 25, no. 6, 2020.

[5] NS-3 Network Simulator Manual, NS-3 Consortium, 2021.

[6] R. Skaggs-Schellenberg, N. Wang, D. Wright, "Performance Evaluation and Analysis of Proactive and Reactive MANET Protocols at Varied Speeds," 10th Annual Computing and Communication Workshop and Conference (CCWC), pp. 0981-0985, 2020.

Graph Representation Learning

[1]*Jakub Ivan VANKO(2nd year),*
Supervisor: [2]*Peter BEDNÁR*

[1,2]Dept. of Cybernetics and Artificial Intelligence, FEEI TU of Košice, Slovak Republic

[1]jakub.ivan.vanko@tuke.sk, [2]peter.bednar@tuke.sk

***Abstract*—This article provides an overview of graph representation learning, which involves creating low-dimensional representations of complex systems represented as graphs. We cover three main groups of graph representation learning methods: factorization-based, random walk-based, and deep learning-based methods, discussing the advantages and limitations of each approach. In this paper, we also present the current state of our research focused on knowledge graphs in various domains and the use of graph representation learning for prediction tasks.**

***Keywords*—embeddings, graph data, graph representation learning, machine learning**

I. Introduction

Graphs are both a data structure and a universal language used to describe complicated systems. It is a mathematical model for representing the connections between objects. In addition to being widely utilized in computer science to represent networks like social networks and the Internet, these structures are also frequently used in other disciplines, including biology, physics, and linguistics. At their most fundamental level, graphs are a collection of objects (nodes) and interactions (edges) between pairs of these nodes [1]. Nodes, also known as vertices, are fundamental building components of a graph; they represent entities (objects) in the graph, while edges are the connections (links), which reflect the relationship between the vertices. Both nodes and edges can be given attributes, which are extra pieces of information or characteristics. Node attributes may include information like the node's name, label, weight, or any other relevant data or metadata. For instance, in a social network graph, a person's node attribute may be their name, age, or address. Edge attributes can include information such as the weight of an edge (representing the strength of the connection between two nodes), the edge's direction, or any related data or metadata. If we take again as an example a social network graph, the edge attribute between two people could be the type of relationship between them—whether they are friends, relatives, or colleagues.

II. Machine learning on graphs

The standard procedure of machine learning on graphs typically starts with creating a meaningful representation for the objects of interest (nodes, edges, or whole graphs), and then using that representation to train a predictor for the intended task. We want to mathematically constrain these objects so that similar objects are mathematically close. However, it might be challenging to characterize this similarity precisely in graph machine learning because, for example, two nodes could be more similar when they have the same labels or the same neighbors. Generally, there are 3 main groups of tasks in machine learning problems that can be solved on graphs - *Node-level tasks* (most often solves the problem of node classification) [2], *Edge-level tasks* (solves problem of relation prediction) [3], and *Graph-level tasks* (tasks like graph generation and graph classification). The traditional machine learning approach is about designing right features for nodes, links, or graphs; the process of designing features is called *feature engineering*. However, this process is time-consuming and takes most of the effort. The most efficient way would be to do this process automatically, meaning we would learn the features and structure of the graph we are interested in. This automation is called *representation learning* – efficient task-independent feature learning for machine learning with graphs [4].

Graph representation learning focuses on learning meaningful and low-dimensional representations of graph data. The idea behind representation learning is that we map the graph (or parts of the graph) into d-dimensional space and represent it as a vector of d numbers. This vector is also called a feature representation or an embedding, and it can be used for machine learning tasks. The goal of an embedding is to preserve the structural information of the original space in the embedded space in a way that makes it easier to perform mathematical operations and analysis. There are a variety of methods for graph representation learning, each with its own strengths and weaknesses. We divide these methods into 3 main groups - *factorization-based methods*, *random-walk based methods* and *deep learning-based methods*.

A. Factorization-based Methods

Factorization-based methods factorize the matrix, which represents the connections between nodes (i.e., the adjacency matrix), to generate the embedding. It tries to represent each node's embedding as a linear combination of its neighbors' embeddings [5]. Their biggest advantage is that they are simple to implement and interpret. They are computationally efficient and can handle large-scale graph data. The ability to capture complex graph structures and relationships is their biggest limitation. The factorization can be performed using various techniques, including *Singular Value Decomposition (SVD)*, or *Non-negative Matrix Factorization (NMF)*.

SVD is a mathematical technique used to break down a matrix into three components: left singular vectors, right singular vectors, and singular values. These components provide a way to understand the underlying structure of the matrix and identify the most important information in it. The left and

right singular vectors are matrices that represent the direction of maximum variance in the row and column spaces of the matrix, respectively. The singular values are numbers that represent the amount of variance in the data that is captured by each singular vector. By decomposing a matrix using SVD, we can identify the most important features of the data and represent it in a more compact form [6].

Non-negative Matrix Factorization (NMF) is a technique used to break down a non-negative matrix into two non-negative matrices of lower rank. We approximate the original matrix as a product of two matrices, one representing the basis vectors and the other representing the coefficients that combine the basis vectors. The advantage of NMF is that it produces a sparse factorization (many of the coefficients in the matrix are zero), which can make the factorization more interpretable and can help identify the most important features in the data.

B. Random Walk-based Methods

Random walk-based methods simulate a random walk on a graph, where the walker randomly traverses the nodes and edges of the graph. By repeating this process multiple times and collecting the resulting sequences of visited nodes, these methods construct a neighborhood matrix or transition matrix that can be used to derive node embeddings. The neighborhood matrix captures the local structural information of the graph, where each entry represents the number of times a node was visited in the context of another node. The transition matrix, captures the global structural information of the graph, where each entry represents the probability of transitioning from one node to another. These two matrices are then used to perform matrix factorization or other dimensionality reduction techniques to obtain embeddings of the nodes in the graph. Some of the well-known random walk-based methods are DeepWalk and Node2Vec.

Deepwalk was first introduced by Perozzi, Al-Rfou, and Skiena in 2014 [7]. The idea behind DeepWalk is to use a random walk algorithm to generate sequences of nodes in the graph. The node sequences are then used as input to a neural network that learns to predict the likelihood of a node being present in the context of another node. The network is trained using a Skip-gram model (neural network architecture used in natural language processing). Once the network is trained, the node embeddings can be extracted from the learned weights of the hidden layer. These embeddings capture the structural information of the graph.

Node2Vec was introduced by Grover and Leskovec in 2016 [8]. It is very similar to DeepWalk - it generates node sequences for training a Skip-gram model to learn node embeddings. However, Node2Vec uses a biased random walk algorithm to generate the node sequences.The biased random walk algorithm in Node2Vec is based on two parameters: p and q The parameter p controls the likelihood of returning to the previous node, while the parameter q controls the likelihood of exploring a new node. By varying the values of p and q, the algorithm can perform breadth-first sampling, depth-first sampling, or a balance between the two, capturing different types of network structures.

C. Deep Learning-based Methods

The rising popularity of deep learning has led to the use of deep learning based methods on graphs. With their power, it is easier to model non-linear structures, so deep autoencoders have been used for dimensionality reduction.

One of the most common deep learning-based methods for graph representation learning is the graph convolutional network (GCN), which was introduced by Thomas Kipf and Max Welling in 2016 [9]. GCNs are an extension of convolutional neural networks (CNNs) to graph-structured data, and their main goal is to learn a representation for each node and capture its local and global structure within the graph. The main idea behind GCNs is to apply a convolutional operation on the graph data, similar to how CNNs apply convolutional filters to images. In a GCN, the convolution operation involves aggregating the features of each node and its neighboring nodes and then applying a weight matrix to compute a new feature vector for each node. The aggregation step is performed using a message passing scheme, where the feature vectors of neighboring nodes are combined and transformed into a message vector. The message vectors are then combined with the original node feature vectors, and the result is passed through a non-linear activation function (such as the ReLU). The model can capture increasingly complex features of the graph structure by stacking multiple layers of GCNs.

Another popular deep learning-based methods are Graph Attention Networks (GATs), which were introduced in 2018 by Veličković et al. [10]. The main idea behind GATs is to use an attention mechanism to dynamically weigh the contributions of neighboring nodes when computing the new feature vector for a given node. In GAT, the attention mechanism involves computing attention coefficients for each pair of nodes, based on their current feature vectors. The attention coefficients are then used to compute a weighted sum of the feature vectors of the neighboring nodes, which is combined with the original node feature vector to compute a new feature vector. Using this attention mechanism, GAT can learn to assign different weights to different nodes, based on their relevance to the task, allowing the model to capture more nuanced information about the graph structure. GATs also use multi-head attention, which involves computing multiple attention coefficients and feature vectors for each node and then concatenating them together, allowing the model to capture multiple aspects of the graph structure simultaneously.

III. Current progress and future work

In the past year we have focused on extracting and building a knowledge graph from data provided by the Ayanza company [11] from beta version of their application. We successfully extracted and built a heterogenous, dynamic, collaborative knowledge graph, which we designed in last year's article [12]. After building the graph, we conducted several experiments focused on link prediction using graph representation learning. To achieve this goal, we employed various methods to transform the nodes of the extracted graph into embedding representations. Our focus was primarily on random walk-based techniques, including the widely used DeepWalk and Node2Vec algorithms. Given that the knowledge graph contains substantial amounts of textual information, we also experimented with the popular BERT (Bidirectional Encoder Representations from Transformers)[13] language model. Using the knowledge graph in embedding representation as input, we evaluated the performance of various popular link

prediction models, including TransE [14], ConvE [15], and autoKGE [16]. Across all of our experiments, the mean reciprocal rank (MRR) was always up to 0.69, with the best performing model being autoKGE. Upon further review of our experimental results, we identified the sparsity of our graph due to the limited number of users in the beta version as a significant factor in our model's performance. Since then, we have been granted access to the production database of Ayanza, from which we can extract a much denser graph. Besides the data provided by Ayanza, we are currently also creating a knowledge graph of botanical data, which we are extracting from WikiData for application called Tomappo [17]. In our further research, we will focus on deep neural methods for graph prediction that can embed both nodes and links. Our goal is to design a deep neural network for graph prediction that can be used regardless of the domain specification.

IV. Conclusion

The field of graph representation learning has made significant progress in recent years, with a wide range of methods and algorithms proposed to learn effective representations of graph-structured data. From more traditional methods like Singular Value Decomposition and Non-Negative Matrix Factorization to more recent techniques like DeepWalk, Node2Vec, GCNs, and GATs, each approach offers unique advantages in terms of scalability, interpretability, and performance. In our further research, we will focus on experiments using graph representation learning methods for prediction tasks on various graph data.

References

[1] W. L. Hamilton, R. Ying, and J. Leskovec, "Representation learning on graphs: Methods and applications," *CoRR*, vol. abs/1709.05584, 2017. [Online]. Available: http://arxiv.org/abs/1709.05584

[2] P. Kazienko and T. Kajdanowicz, "Label-dependent node classification in the network," *Neurocomputing*, vol. 75, pp. 199–209, 01 2012.

[3] M. Wang, L. Qiu, and X. Wang, "A survey on knowledge graph embeddings for link prediction," *Symmetry*, vol. 13, no. 3, p. 485, Mar 2021. [Online]. Available: http://dx.doi.org/10.3390/sym13030485

[4] J. Leskovec, "Cs224w: Machine learning with graphs," 2021.

[5] X. Zhang, K. Xie, S. Wang, and Z. Huang, "Learning based proximity matrix factorization for node embedding," 2021. [Online]. Available: https://arxiv.org/abs/2106.05476

[6] A. Fonarev, "Matrix factorization methods for training embeddings in selected machine learning problems," Ph.D. dissertation, Skolkovo Institute of Science and Technology, 2018.

[7] B. Perozzi, R. Al-Rfou, and S. Skiena, "Deepwalk: Online learning of social representations," *CoRR*, vol. abs/1403.6652, 2014. [Online]. Available: http://arxiv.org/abs/1403.6652

[8] A. Grover and J. Leskovec, "node2vec: Scalable feature learning for networks," *CoRR*, vol. abs/1607.00653, 2016. [Online]. Available: http://arxiv.org/abs/1607.00653

[9] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *CoRR*, vol. abs/1609.02907, 2016. [Online]. Available: http://arxiv.org/abs/1609.02907

[10] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," 2017. [Online]. Available: https://arxiv.org/abs/1710.10903

[11] [Online]. Available: https://ayanza.com/

[12] J. I. Vanko, "Link prediction in knowledge graphs," *22nd Scientific Conference of Young Researchers*, p. 52–55, 2022.

[13] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," 2019.

[14] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko, "Translating embeddings for modeling multi-relational data," in *Advances in Neural Information Processing Systems*, C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Weinberger, Eds., vol. 26. Curran Associates, Inc., 2013. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2013/file/1cecc7a77928ca8133fa24680a88d2f9-Paper.pdf

[15] T. Dettmers, P. Minervini, P. Stenetorp, and S. Riedel, "Convolutional 2d knowledge graph embeddings," *CoRR*, vol. abs/1707.01476, 2017. [Online]. Available: http://arxiv.org/abs/1707.01476

[16] Y. Zhang, Q. Yao, W. Dai, and L. Chen, "Autosf: Searching scoring functions for knowledge graph embedding," 04 2020.

[17] "Tomappo." [Online]. Available: https://tomappo.com/

Deep Learning for Upper Limb Rehabilitation

[1]*Stanislav HUSÁR (2nd year),*
Supervisor: [2]*Marek BUNDZEL*

[1,2]Department of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1]stanislav.husar@tuke.sk, [2]marek.bundzel@tuke.sk

***Abstract*—This paper describes recent advances in developing our upper limb rehabilitation solution. We are focusing on the diagnostics and rehabilitation of patients suffering from tremors. We are developing the Rehapiano device, which uses load cells to measure forces exerted by the patient's hands. We describe various deep neural network architectures that combine convolutional and LSTM layers to classify spectrogram images of input signals. We propose to use this pipeline to classify the type and severity of tremors and their underlying diagnosis.**

***Keywords*—intelligent rehabilitation, tremor, load cell, short-time Fourier transform, spectrogram, CNN, LSTM**

I. Introduction

Many people worldwide suffer from various motoric impairments caused by diseases and injuries. These significantly reduce their quality of life, making daily activities difficult. A traditional way to heal these impairments is rehabilitation consisting of endless repetitions of simple movements. Unfortunately, patients' willingness to perform these is declining, and rehabilitation often fails. Furthermore, diagnosing these impairments is very demanding for a physician who thoroughly examines the patient's movements.

Advances in technology enable the gamification of rehabilitation. Gamification hides performed movements behind some goal that is funny to achieve. It helps to increase and sustain patients' willingness to perform movements. Physician's burden is also significantly decreased. He can repeatedly examine all recordings for a more precise diagnosis.

The use of rehabilitation devices also enables artificial intelligence processing of recordings. This approach can provide better insight into the patient's diagnosis, severity, and overall state. It can also adapt the rehabilitation game's difficulty and other properties to the patient's needs.

Our research focuses on developing upper limb rehabilitation and diagnostics platform focused around a novel device, the Rehapiano (Figure 1). The patient will insert fingers into their respective posts and use them to exert force on load cells. This force will control the rehabilitation game and serve as input data to deep learning models.

We have validated the ability of Rehapiano to detect frequencies of interest. We used a DC motor with an eccentric load to generate controlled vibrations. In addition, we used an incremental encoder and closed-loop control algorithm to stabilize vibrations at specific frequencies.

This publication was realized with support of the Operational Programme Integrated Infrastructure in frame of the project: Intelligent systems for UAV real-time operation and data processing, code ITMS2014+: 313011V422 and co-financed by the European Regional Development Found

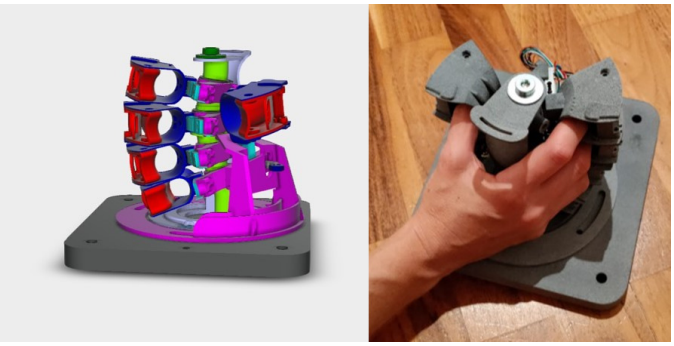


Fig. 1. Rehapiano device CAD drawing and photograph

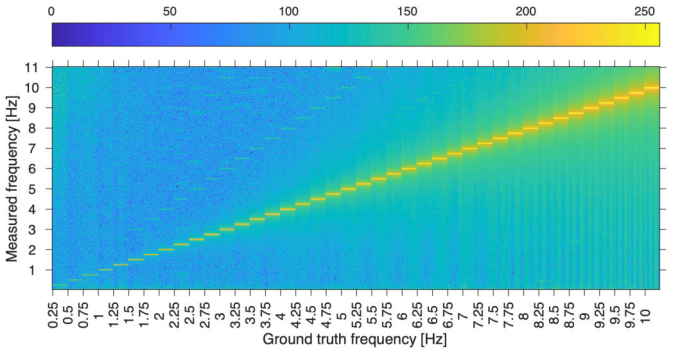


Fig. 2. Rehapiano device frequency response

Figure 2 shows the power spectra of all validated frequencies. Here we can see that the Rehapiano can measure all frequencies of interest. The amplitude of vibrations is proportional to their frequency. DC motor dynamics mainly cause this. We regulate frequency by changing motor power, inevitably changing the amplitude of vibrations.

II. Problem Description

We designed the Rehapiano device to diagnose and rehabilitate patients suffering from upper limb tremors. A tremor is an involuntary oscillatory motion. Depending on the patient's diagnosis and condition, it can manifest during rest or various activities. However, all tremors reduce the ability of the patient to perform coordinated movements and, ultimately, any daily activities.

As a diagnostics device, Rehapiano has the potential to help physicians determine the proper diagnosis of the patient and its severity.

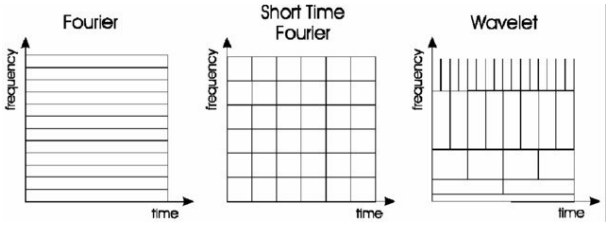


Fig. 3. Frequency and Time-Frequency representations [1]

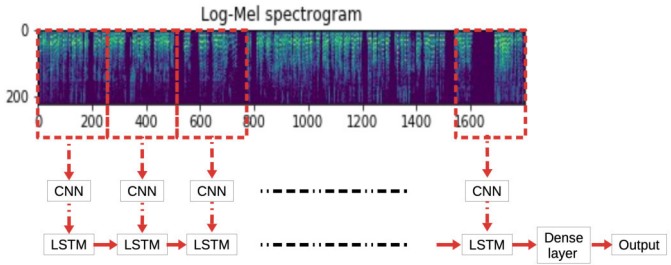


Fig. 4. Basic CNN-LSTM network architecture [5]

III. Solution design

Measurements from sensors come in time representation (as a time series). In time representation, measurements have an excellent temporal resolution but lack frequency information. We can transform data to a frequency representation using Fourier Transform (Figure 3 left) [2]. On the other hand, in frequency representation, measurements have excellent frequency resolution but lack temporal resolution. A simple solution is to use Short Time Fourier Transform, sliding a window over the signal, and computing the Fourier Transform of separate windows (Figure 3 centre). It creates a Time-Frequency representation, giving both temporal and frequency resolution. However, there is a tradeoff situation between temporal and frequency resolution. Another exciting transformation, which does not have this disadvantage, is Wavelet Transform (Figure 3 right). It decomposes a signal into a mother wavelet function combined with its scaled and translated copies.

The convolutional layer is an essential component of deep neural networks [3]. Convolutional layers consist of kernels that learn significant features on the input image. The convolutional layer output is a feature map containing information on where these features were detected. We can create an excellent local feature extractor by stacking multiple convolutional layers in series. However, they cannot detect global features across the entire input.

In this, recurrent LSTM layers are helpful [4]. LSTM layers contain cells containing hidden states and gates operating on them. These hidden states help the layer detect significant relations between distant data portions.

The basic architecture of a deep neural network that combines Time-Frequency representations with convolutional and LSTM layers is in Figure 4[5]. First, the authors transformed input data into a Time-Frequency representation (essentially a 2D image). The authors used logarithmic and Mel scales to create this image. In general linear scale can be used. Next, the authors passed slices of this image through convolutional layers. Then they passed the resulting sequence through LSTM layers. Finally, they used a dense layer for the final classification.

Jianyou Wang et al. have mitigated the time versus frequency resolution tradeoff using two convolutional layers with a different balance of resolutions and a dual-sequence LSTM layer [6]. Soo Hyun Bae et al. instead used a parallel combination of convolutional and LSTM layers [7].

We propose to use architecture similar to figure 4 to classify tremor severity and likely diagnosis. First, we will transform data from individual sensors using Short-Time Fourier Transform and stack them over the third dimension. This transform produces a 10-channel spectrogram image. Then we can use convolutional layers, which classify virtually all images, including spectrograms of periodic signals (like tremors).

A minor disadvantage of convolutional layers is that we must slice signals of varying lengths to produce spectrogram images with fixed dimensions. It can be mitigated by appending recurrent layers. They have proven excellent extractors of relations between distant parts of the input sequence. Recurrent layers conclude the feature extraction stage. Dense layers will produce the final classification of the patient's diagnosis and severity.

IV. Future work

We have completed the Rehapiano hardware design and are testing it to ensure fluent error-free operation in a clinical environment. Meanwhile, we are finalizing software that will allow clinicians to operate the device and enable the collection of the dataset.

Soon, we plan installation in the clinical environment for dataset collection. Dataset collection is a crucial step for future research on the intelligent classification of patient status and adapting the rehabilitation game.

After this, we plan to train and validate described deep learning models on collected data..

References

[1] E. Rinta-Runsala, "Drive system monitoring: Requirements and suggestions," *VTT Case Study Report, Finland*, 2000.

[2] M. K. Kıymık, İ. Güler, A. Dizibüyük, and M. Akın, "Comparison of stft and wavelet transform methods in determining epileptic seizure activity in eeg signals for real-time application," *Computers in biology and medicine*, vol. 35, no. 7, pp. 603–616, 2005.

[3] D. Huang and B. Chen, "Surface emg decoding for hand gestures based on spectrogram and cnn-lstm," in *2019 2nd China Symposium on Cognitive Computing and Hybrid Intelligence (CCHI)*. IEEE, 2019, pp. 123–126.

[4] J. Zhu, H. Chen, and W. Ye, "A hybrid cnn–lstm network for the classification of human activities based on micro-doppler radar," *Ieee Access*, vol. 8, pp. 24 713–24 720, 2020.

[5] A. Meghanani, C. Anoop, and A. Ramakrishnan, "An exploration of log-mel spectrogram and mfcc features for alzheimer's dementia recognition from spontaneous speech," in *2021 IEEE Spoken Language Technology Workshop (SLT)*. IEEE, 2021, pp. 670–677.

[6] J. Wang, M. Xue, R. Culhane, E. Diao, J. Ding, and V. Tarokh, "Speech emotion recognition with dual-sequence lstm architecture," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 6474–6478.

[7] S. H. Bae, I. K. Choi, and N. S. Kim, "Acoustic scene classification using parallel combination of lstm and cnn." in *DCASE*, 2016, pp. 11–15.

Experimental Verification of High Frequency Signal Injection Sensorless Control of PMSM

[1]*Viktor Petro (3rd year),*
Supervisor: [2]*Karol Kyslan*

[1,2]Dept. of Electrical Engineering and Mechatronics, FEI TU of Košice, Slovak Republic

[1]viktor.petro@tuke.sk, [2]karol.kyslan@tuke.sk

***Abstract*—One of the most recent high-frequency (HF) signal injection method is the HF pulse signal injection. It is a simple but effective approach for sensorless control of the PMSM. Control sequences are introduced and one sequence consists of three pulse width modulation (PWM) periods. One PWM period is used for the field-oriented control (FOC) and the rest are used for HF voltage pulse injection into the observed d-axis voltage reference. In this paper an implementation and experimental verification of the HF pulse signal injection-based sensorless control is presented.**

***Keywords*—Sensorless control, high-frequency signal injection, permanent magnet synchronous machine, position estimation**

I. Introduction

Permanent magnet synchronous motor (PMSM) is preferred in the industry due to its high torque density, efficiency, and maintenance-free operation. The field-oriented control (FOC) became a standard control strategy for precise motion control over the full speed range. FOC for PMSM is usually based on the position of the rotor magnetic field. Therefore, the knowledge of the exact rotor electrical angle is essential. Usually, mechanical sensors such as the incremental rotary encoder (IRC) or resolver are used. However, the trend is to eliminate the position sensor because of the electromagnetic noise sensitivity, unreliability, and extra space requirement for mounting purposes. Control approaches that eliminate the mechanical position sensor are referred to as sensorless control methods.

In general, these methods are applicable either for the high-speed or low-speed regions. Methods used for the high-speed regions are usually based on the back electromotive force (back-EMF) estimation. These methods are not usable for low-speed operation. On the other hand, high-frequency (HF) signal injection methods have been successfully applied for low-speed sensorless control of the PMSM. HF pulse signal injection method is one of the most recent methods for HF signal injection. This article will present an experimental implementation and verification of the HF pulse signal injection using the OPAL-RT hardware-in-the-loop (HIL) real-time simulator.

II. HF pulse signal injection

The mathematical background of the HF pulse signal injection is described in [1]. The main idea of this approach is to separate the FOC and the HF injection. It is achieved by creating sequences. One sequence consists of three PWM periods and in our application, the length of it was 300 µs as can be seen in Fig. 1. In the first period within the sequence the algorithm of the FOC is executed. It is followed by two injection periods in which a positive and a negative voltages are injected into the observed real axis (observed d-axis). The currents are sampled at the beginning of each period. The samples at the beginning and at the end of the injection periods are compared resulting in the current differences. The HF injection is performed only in the real axis while the voltage in the imaginary axis is held at zero during the injection periods. Therefore, if the electrical rotor position is correct, the current differences on the imaginary axes will have zero value. The current differences of the imaginary axes are serving as an input to the phase-locked-loop (PLL). PLL is used for electrical position estimation as shown in Fig. 2.

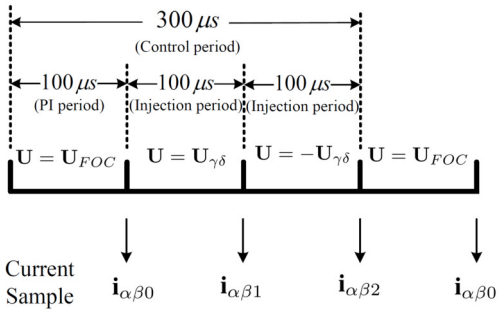


Fig. 1. A sequence of the FOC and HF injection periods, where $\gamma\delta$ represents that the injection is performed in the observed rotor reference frame

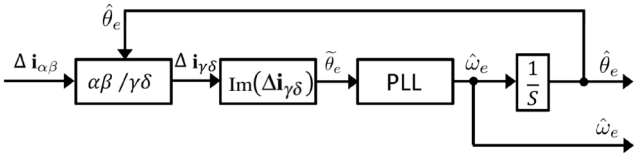


Fig. 2. The coordinate transformation and PLL used for position and rotor speed observation during HF pulse signal injection [2].

In Fig. 2, $\Delta\bar{i}_{\alpha\beta}$ represents the current derivatives in $\alpha\beta$ reference frame, $\Delta\bar{i}_{\gamma\delta}$ are the transformed current derivatives into the estimated rotor reference frame, $\hat{\theta}_e$ is the observed electrical angle used for Park transformations, $\tilde{\theta}_e$ is the imaginary part of the current derivatives in the estimated rotor reference frame which is entering the PLL. The output of the PLL is the observed electrical velocity $\hat{\omega}_e$ and its integral is the observed electrical rotor position $\hat{\theta}_e$.

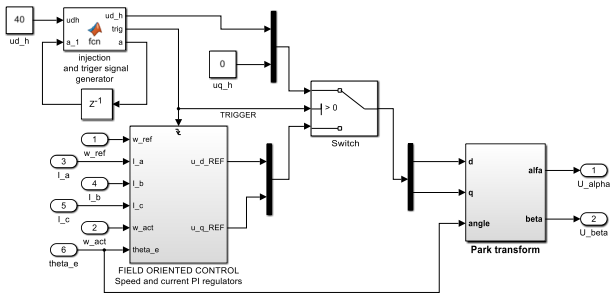


Fig. 3. Block diagram if FOC and HF injection separation in MATLAB/Simulink

III. HF PULSE SIGNAL INJECTION IMPLEMENTATION

The sensorless control based on the HF pulse signal injection was verified using the OP5600 real time simulator. The software called RT-LAB was used. It offers a standard MATLAB/Simulink interface where the execution source code is obtained by automatic code generation from the model which is developed directly in MATLAB/Simulink. Fig. 3 shows the block diagram with separation of the FOC period from the periods with signal injection. The "FIELD ORIENTED CONTROL" block contains the current and speed PI regulators. Normally, the reference voltages from the current regulators are directly entering the Park transformation. For the HF pulse signal injection, a switch is placed between the FOC and the Park transformation block. By proper control of this switch either the FOC or the injection is enabled. The signal entering the Park transformation is shown in Fig. 4. It

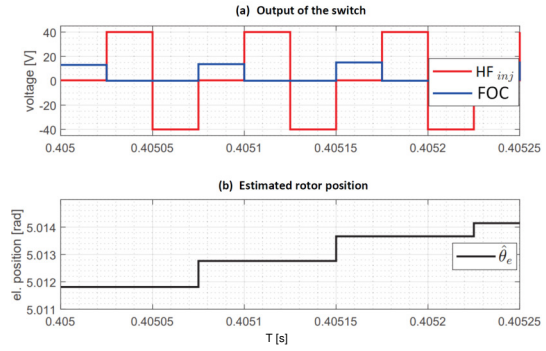


Fig. 4. A sequence of the FOC and HF injection periods.

can be seen in the upper part of Fig. 4, that a suitable voltage (blue), calculated by PI controllers of current, is applied to the PWM modulator. This period is followed by two injection periods where a positive and a negative voltage with the same amplitude is applied. The observed electrical position is constant during one sequence. At the beginning of each sequence the current sample is used as a feedback for the FOC. Then a current sample at the beginning and at the end of the injection periods is ignored by the FOC and these values are used to calculate current derivatives. At the beginning of the new sequence, the current differences are calculated from the previous sequence, the PLL is executed and a new estimated rotor position is observed. With the updated observed electrical position the new current samples are prepared for the next FOC period.

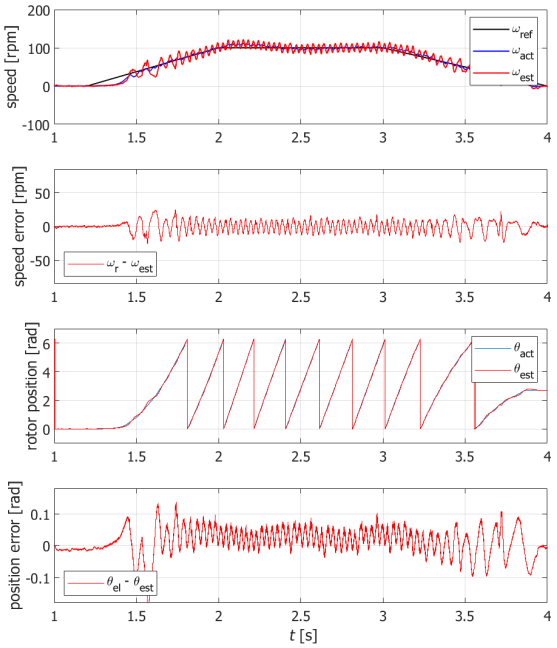


Fig. 5. Experimental results of sensorless control using the HF pulse signal injectrion-based method.

IV. EXPERIMENTAL RESULTS

For the experimental verification, a 2.5 kW interior permanent magnet synchronous motor (IPMSM) was used with the nominal current of 4.2 A, and the inductances $L_d = 22$ mH, and $L_q = 61.5$ mH. The nominal speed is 2300 rpm. In sensorless mode, the reference speed was set to 100 rpm which represents about 4% of the nominal speed. The experimental results are shown in Fig. 5. The observed electrical position was used in Park tranformations and the observed speed was used as a feed-back for speed regulator. The mean observed position error is around 0.05 electrical radians. The observation of speed and electrical rotor position was stable.

V. CONCLUSION

In this article, experimental results of the HF pulse signal injection based sensorless control of IPMSM was presented. The results show the performance of this approach is suitable for speed sensorless control. Further investigation is needed on the amplitude of the injected voltage. Future work will test the observation under loaded conditions.

ACKNOWLEDGMENT

This work was supported by the Faculty of Electrical Engineering and Informatics, Technical University of Košice under Grant FEI-2022-86.

REFERENCES

[1] V. Petro, K. Kyslan, P. Bober, and M. Lacko, "Optimization of injected voltage amplitude for low-speed sensorless control of pmsm with high-frequency pulse signal injection," in *2022 IEEE 20th International Power Electronics and Motion Control Conference (PEMC)*, 2022, pp. 721–727.

[2] G. Xie, K. Lu, S. K. Dwivedi, J. R. Rosholm, and F. Blaabjerg, "Minimum-voltage vector injection method for sensorless control of pmsm for low-speed operations," *IEEE Transactions on Power Electronics*, vol. 31, no. 2, pp. 1785–1794, 2016.

Exploring the correlation between solar cycles and terrestrial phenomena through machine learning analysis

[1]*Martin NGUYEN (1st year),*
Supervisor: [2]*Ján GENČI, Consultant:* [3]*Pavol BOBÍK*

[1,2]Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic
[3]Department of Cosmic Physics, Institute of Experimental Physics SAS Kosice, Slovak Republic

[1]martin.nguyen@student.tuke.sk, [2]jan.genci@tuke.sk, [3]bobik@saske.sk

***Abstract*—Solar cycles have been observed for centuries and are often linked to changes in Earth's climate. It is said that different numbers of sunspots affect the weather patterns, which has a big impact on life on Earth. Even though there are existing researches that study the correlation between solar cycles and various events on the Earth, the relationship is still not fully understood. This study aims to further investigate this relationship by analyzing existing research papers using machine learning and statistical approaches. Our research is focused on the relationship between the solar cycles and the outbreak of wars as well as the disruption and malfunction of electrical grid networks around the world.**

***Keywords*—power grid, machine learning, solar cycle, war correlation**

I. Introduction

Solar cycles refer to the natural fluctuations in the activity of the Sun [1]. Such changes include brightness, magnetic field, and the number of sunspots. Each solar cycle has a different length, but the typical duration of a solar cycle is approximately 11 years. Solar cycles have a range of impacts on Earth, typically it has major effects on climate and weather patterns. Such influence is widely described in the publication [2]. Due to changes in the magnetic field and overall changes in space weather conditions, outages of power grids also often occur [3]. The power outage topic is widely discussed and researched in many research papers, such as [4] where the operational reliability of electric power systems during geomagnetic storms is reviewed. However, a full understanding of the correlation between solar cycles and their impact on Earth still needs to be achieved.

There have been many studies that contributed to this topic over the years while they have been focused on one common goal - to get a deeper understanding of how solar cycles affect the Earth's biosphere and how we can identify potential patterns or trends. In this study, our objective is to investigate the connection between solar activity and life on Earth through the utilization of machine learning and statistical methodologies applied to previously published research articles. Our goal is to gain a deeper understanding of how Earth's biosphere is affected by solar cycles and to identify any patterns that were not found yet.

To achieve this, we will conduct a systematic review of existing data sets from a range of sources such as scientific journals and online databases. The collected data will then be analyzed using a variety of machine learning techniques, such as ARIMA [5], Fourier analysis, and deep learning methods.

The results of this study have the potential to significantly contribute to our understanding of the complex relationship between solar activity and life on Earth and may have important implications for future research in this area.

II. Understanding solar cycle and sunspots

Solar cycles refer to the periodic variation of different forms of solar activity, such as the number of sunspots, flares, or mass ejections which occur directly on the surface of the Sun. Scientists have been studying solar cycles for over a century and one of the prevailing theories is that solar cycles are caused by the interaction between the Sun's magnetic field and the flow of plasma. This interaction then transforms the Sun's magnetic field which leads to various solar activities and the formation of sunspots, depending on the strength and orientation of the magnetic field. An example of modern observations is Solar Dynamics Observatory (SDO) NASA mission [6], where the Sun is being observed since 2010. This mission is part of the Living With a Star program [7].

The solar cycle has a period of approximately 11 years, but the length can vary between 9 and 14 years, for example, one of the shortest solar cycles was between the years 1766 and 1775 and the longest was between the years 1784 and 1798.

Each solar cycle is characterized by the increase of solar activity, solar maximum, decrease of solar activity, and solar minimum, which is then followed by another solar cycle. The history of solar cycles is shown in figure 1. The solar maximum is not always reached in the middle of the cycle.

There are several practical applications for understanding solar cycles. For example, during solar maxima, electrical devices are more often damaged than during solar minima. For example, when rates and strength of solar flares are increased, the satellites are prone to be damaged or electrical currents on power lines are affected on power grids which can cause power outages.

There are also periods with unusual low and high solar activity, for example, Maunder Minimum which occurred between the years 1645-1715, the Dalton maximum which occurred between the years 1790-1820, and the Modern maximum

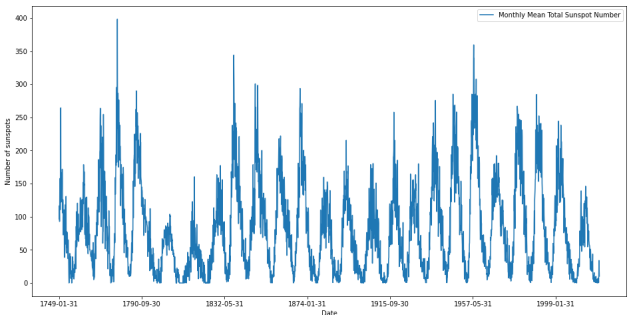


Fig. 1. Number of sunspots throughout the years 1755-2019

which occurred between the years 1933-2008. In our study, we research if these minimum and maximum periods have an impact on various events.

III. Analysis of Correlation Between Solar Cycles and Climate Change

The Sun is a key driver of Earth's climate and its activity strongly influences climate change. Previous studies suggested a link between solar cycles and climate, such as varying temperatures due to changing CO_2 concentration [8]. Therefore we also aim to build on that research to gain a deeper understanding of this relationship.

The authors in [9] compared the cycle length series to global and northern hemisphere temperatures from different temperature series. They found that all the series show similarities in the period of 1870-1970 and deviate before and after that period. Both global and northern hemisphere temperatures increase sharply after 1979 while the cycle length series decreased. Overall they found that there is a poor correlation between the cycle length and temperature series. The authors concluded that the previously reported correlation between the cycle length series and temperatures on the Earth was an artifact and that the warming influence on Earth is predominantly due to increases in atmospheric greenhouse gas concentrations since 1750.

IV. Analysis of the Correlation Between Solar Cycles and Space Weather

It is known that there is a specific correlation between space weather and solar cycles, for example, the more frequent occurrence of geomagnetic storms during solar minimum and maximum and global disturbances in Earth's magnetic field. Some studies test hypotheses about the occurrence of geomagnetic storms during the solar cycle. One example is a study [10] that uses a 150-year record of global geomagnetic activity and probabilistic models to test these hypotheses. The results displayed in figure 2 show that storms occur more frequently during the active phase around solar maximum and the most extreme events are likely to occur more frequently during longer solar cycles. The study also finds that extreme events occur earlier in even-numbered cycles and later in odd-numbered cycles. We can use these findings to estimate the probability of extreme event occurrence in future solar cycles, which emphasizes the importance of solar cycle prediction to be more prepared for activities affected by extreme space weather by planning and scheduling.

Another study [11] investigates the FORMOSAT-3 (FS3) satellite under different space weather conditions. The results

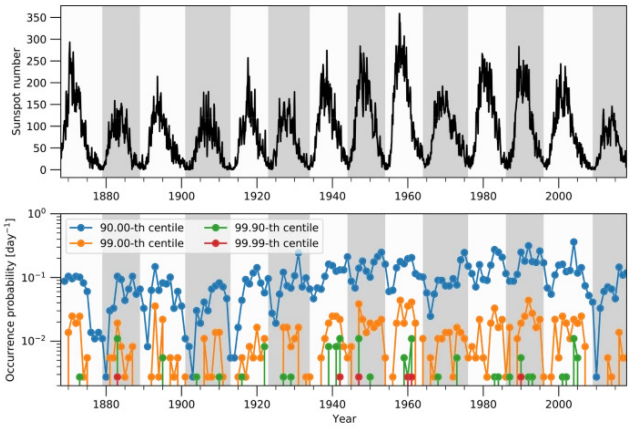


Fig. 2. Top: Monthly sunspot number, with even-numbered solar cycles shaded grey. Bottom: The occurrence probability of geomagnetic storms using different thresholds for storm magnitude.
Image from: [10].

showed that more negative anomalies have a negative correlation with the level of geomagnetic activity and sunspot numbers during different parts of the solar cycle. The study found that the anomalies mostly occurred during low solar activity and they were most likely caused by high-energy trapped protons and cosmic rays. Based on that finding, satellites are prone to suffer during low solar or geomagnetic activity. Previous studies have suggested that satellite anomalies are often related to charged particles and that many anomalies were attributed to high geomagnetic activity but focus on satellites in geostationary orbit, not low-Earth orbit.

V. Obtaining Data Sets

There is a variety of sources for obtaining data sets, including observational, simulation, and experimental data. Since our study is focused on real-life data, we always try to obtain observational data first. At the beginning of our study, we tried to collect direct measurements of solar cycles, which were measured by astronomical observatories. Such data sets contain observational data for the number of sunspots, date, solar flare counts, or measurements of the solar magnetic field. These data sets are publicly available on public databases and APIs.

Since we focus on the correlation between power outages, war conflicts, and solar activity, our other field of interest are data sets about power grid outages and world conflicts. Such data sets were collected with a web scraping approach from public sites or downloaded from other research or data set sites like the IEEE (Institute of Electrical and Electronics Engineers) DataPort site. One example is the electricity grid outage data set [12] which is a simulation of an electricity network in Western Sierra Leone. The data set contains many useful information that we can use in our research. It comprises 10 generators, 50 buses, and 40 load zones serving 270,000 customers and it includes comma-separated values files containing information on network components, generators, load profiles, component failure modes, repair time distribution, and maintenance team mapping. By combining similar data sets and after preprocessing, we can use them directly in our research.

It is always important to preprocess and clean the data to ensure its quality, correctness, and completeness once we

obtain the data. In solar cycle data sets, we filtered down date intervals, where complete solar cycles are available. We also added the information about Maunder minimum, Dalton maximum, and Modern maximum, we marked the maximum of the solar cycle and computed solar cycle duration.

VI. Use of machine learning methods on forecasting solar cycles

Machine learning methods have been increasingly used in recent years to study solar cycle events, including solar flares, coronal mass ejections, and sunspot activity. These methods can be used to improve our understanding of these phenomena and to make predictions about future solar activity.

One of the main advantages of using machine learning methods is their ability to analyze large amounts of data and to identify patterns and relationships that are not immediately apparent to a human. For example, using machine learning methods, researchers can analyze data from multiple instruments and observatories to identify patterns in solar activity that might not be visible in data from a single instrument.

One of the machine learning methods used for forecasting solar cycle events, such as predicting the number of sunspots, is ARMA. In the [13], authors used the method and made predictions based on the data set of historical records from the 18th to 21st century. The proposed result has seemed correct as is displayed in figure 3. We experimented with the proposed ARMA model for forecasting by executing the exact model as described in the publication on the test database. We then forecasted the following solar cycle and got non-satisfactory results as it is shown in figure 4.

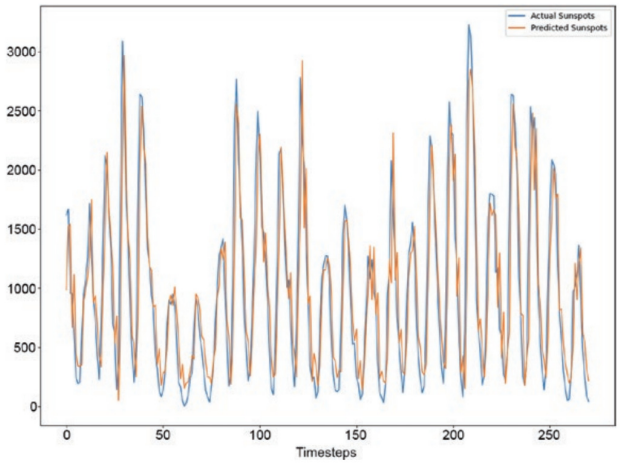


Fig. 3. Actual and fitted values of the ARMA model

Other publications [14] that use statistical models like ARIMA or AR relies on the correlation between their predictions and NOAA (National Oceanic and Atmospheric Administration) predictions. In their work, they show that the predictions made by the model and the predictions made by NOAA on the estimated curve of monthly mean predicted sunspot numbers during the years 2018 and 2019 are consistent and that they are in good agreement. The correlation coefficient between the predictions made by the model and NOAA's predictions were found to be 96.5%, indicating a strong correlation.

In the paper [15], the authors conducted a comparative study of non-deep learning and deep learning models as well as

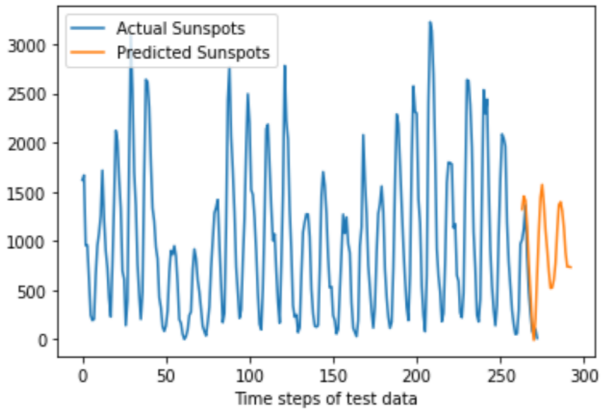


Fig. 4. Forecast made with ARMA model

their ensemble models for sunspot number prediction. Three non-deep learning models SARIMA, Exponential Smoothing, and Prophet, four deep learning models (LSTM, GRU, Transformer, and Informer) are compared, as well as five ensemble models (mean, median, error-based method, linear regression, and XGBoost). The study found that the XGBoost-DL model, which uses XGBoost to ensemble the four deep learning models, has the best performance in comparison with other models and the predictions of the National Aeronautics and Space Administration (NASA). The study also proposed an open-source Python package of the XGBoost-DL model for sunspot number prediction and used it to forecast Solar Cycles 25 and 26, comparing the results with NASA's prediction.

Good results also propose authors in [16] where the LSTM+ model was proposed as the optimal version of an LSTM model to predict solar activity using F10.7 data and SSN data from 1954 to 2019. Their model's parameters were fine-tuned for improved prediction accuracy. It was validated with solar cycle 23 and 24 data, showing better results compared to the backpropagation model. The long-term prediction for solar cycle 25 was made with an average error of 6.6% supported by a preliminary forecast from NOAA/NASA.

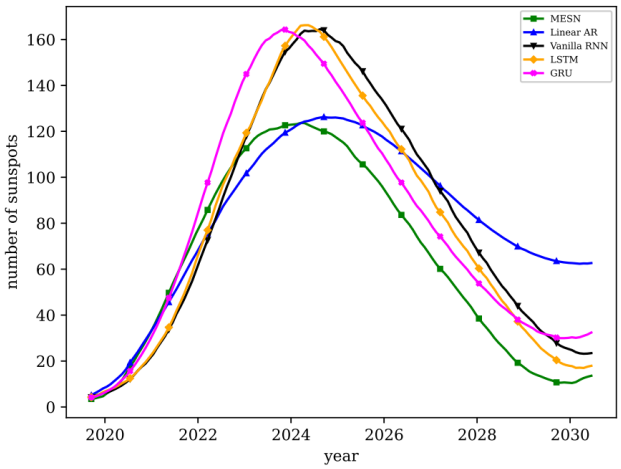


Fig. 5. Comparison of different forecasting algorithms for Solar Cycle 25. Image from: [17].

However, it still looks like even with a neural network approach, only the following solar cycle can be almost precisely predicted. The authors in [17] used five Recurrent Neural Network algorithms to forecast Solar Cycle 25. Such a comparision is shown in figure 5. The best performance

was found with the Echo State Network (ESN) but even this algorithm was limited to forecasting only one cycle ahead. The authors then designed a Modified ESN (MESN) algorithm which produced more accurate and robust forecasts. The best forecasts were obtained when the training data was from the rising phase of the cycle. Both MESN and ESN agreed that Cycle 25 will last for about 10 years and the peak will occur in April 2024 with a maximum number of sunspots forecasted to be 137. Both algorithms indicated that Cycle 25 will be slightly stronger than Cycle 24 and weaker than Cycle 23. The authors used both simulated data from a physics-inspired solar dynamo model and observational data to make their forecasts.

VII. Use of machine learning methods on finding correlation with solar cycles

Scientists have been trying to find correlations between solar cycles and other phenomena for many years. Solar cycles, which are characterized by regular changes in the number of sunspots, have been shown to have a significant impact on the Earth's climate and environment. Some researchers are trying to find the correlation between solar-related events, such as solar irradiance and terrestrial insolation [18]. Other researches focus on the impact on satellites [19], GPS navigations [20] and power grids [21]. Some research papers even try to find correlations between the number of publications of specific authors, as is described in [22].

In recent years, the use of Machine Learning methods has also been used to study the correlation between solar cycles and other phenomena. The ability of these methods to analyze large amounts of data and identify patterns and relationships that might not be immediately apparent to human analysts has proven to be valuable in this research area. Such example is research about the correlation with sunspot groups and coronal mass ejections by determining source active regions using neural networks [23].

In another research [24], authors trained an SVM model to recognize patterns in the magnetic fields of active regions on the sun, which are areas of intense magnetic activity. These patterns can indicate whether a flare is likely to happen. The model looked at various features of the magnetic fields, such as the amount of energy stored in them, and how they change over time. The model was able to successfully predict flares with an accuracy of around 75-89%. The study found that certain features of the magnetic fields are important for predicting flares than others and that the amount of energy stored in the magnetic fields increased significantly before a flare occurred.

VIII. Conclusion and future work

The study of the relationship between solar cycles and other aspects of the environment is a dynamic and rapidly evolving field. Our study provides a comprehensive overview of current studies in this field and serves as a basis for future research.

In our future work, we will investigate the correlation between power outages, war cycles, and solar cycles. The availability of data on power outages, solar activity as well as other measurable data provides a valuable opportunity to explore the connection between these factors. This research could have important implications for the reliability and predictability of electrical systems in the future and could provide insights into ways to mitigate the impact of solar activity on power grids and other aspects of life on Earth.

IX. Acknowledgment

This publication was realized with the support of the Operational Programme Integrated Infrastructure in the frame of the project: Intelligent systems for UAV real-time operation and data processing, code ITMS2014+: 313011V422, and co-financed by the Europen Regional Development Found.

References

[1] A. Balogh, H. S. Hudson, K. Petrovay, and R. von Steiger, "Introduction to the solar activity cycle: Overview of causes and consequences," *Space Sciences Series of ISSI*, p. 1–15, 2015.

[2] J. D. Haigh and P. Cargill, *The Sun's influence on climate*. Princeton University Press, 2015.

[3] *Impacts of severe space weather on the electric grid*. Mitre Corp., JASON Program Office, 2011.

[4] J. G. Kappenman, "An introduction to power grid impacts and vulnerabilities from space weather," *Space Storms and Space Weather Hazards*, p. 335–361, 2001.

[5] S. Ho and M. Xie, "The use of arima models for reliability forecasting and analysis," *Computers & Industrial Engineering*, vol. 35, no. 1-2, p. 213–216, 1998.

[6] W. D. Pesnell, B. J. Thompson, and P. C. Chamberlin, "The solar dynamics observatory (sdo)," *Solar Physics*, vol. 275, no. 1-2, p. 3–15, 2011.

[7] G. Withbroe, M. Guhathakurta, and J. Hoeksema, "Origins of the international living with a star program," *Advances in Space Research*, vol. 35, no. 1, p. 40–43, 2005.

[8] R. Booth, "On the influence of solar cycle lengths and carbon dioxide on global temperatures," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 173, p. 96–108, 2018.

[9] T. Chatzistergos, "Is there a link between the length of the solar cycle and earth's temperature?" *Rendiconti Lincei. Scienze Fisiche e Naturali*, 2022.

[10] M. J. Owens, M. Lockwood, L. A. Barnard, C. J. Scott, C. Haines, and A. Macneil, "Extreme space-weather events and the solar cycle," *Solar Physics*, vol. 296, no. 5, 2021.

[11] H.-W. Shen, J.-H. Shue, J. Dombeck, and T.-P. Lee, "An evaluation of space weather conditions for formosat-3 satellite anomalies," *Earth, Planets and Space*, vol. 73, no. 1, 2021.

[12] H. George-Williams, N. Kebir, S. Hirmer, and M. McCulloch, "Sample electricity grid outage management data," 2021. [Online]. Available: https://dx.doi.org/10.21227/cwkn-9888

[13] J. Korstanje, *Advanced forecasting with python: With state-of-the-art-models including LSTMs, Facebook's prophet, and Amazon's DeepAR*. Apress, 2021.

[14] H. Abdel-Rahman and B. Marzouk, "Statistical method to predict the sunspots number," *NRIAG Journal of Astronomy and Geophysics*, vol. 7, no. 2, p. 175–179, 2018.

[15] Y. Dang, Z. Chen, H. Li, and H. Shu, "A comparative study of non-deep learning, deep learning, and ensemble learning methods for sunspot number prediction," *Applied Artificial Intelligence*, vol. 36, no. 1, 2022.

[16] H. Zhu, W. Zhu, and M. He, "Solar cycle 25 prediction using an optimized long short-term memory mode with f10.7," *Solar Physics*, vol. 297, no. 12, 2022.

[17] A. Espuña Fontcuberta, A. Ghosh, S. Chatterjee, D. Mitra, and D. Nandy, "Forecasting solar cycle 25 with physical model-validated recurrent neural networks," *Solar Physics*, vol. 298, no. 1, 2023.

[18] A. Hempelmann and W. Weber, "Correlation between the sunspot number, the total solar irradiance, and the terrestrial insolation," *Solar Physics*, vol. 277, no. 2, p. 417–430, 2011.

[19] D. Hastings and H. B. Garrett, *Spacecraft–Environment Interactions*. Cambridge University Press, 2004.

[20] B. Hofmann-Wellenhof, H. Lichteneger, and J. Collins, *Global Positioning System: Theory and practice*. Springer, 2005.

[21] D. Boteler, R. Pirjola, and H. Nevanlinna, "The effects of geomagnetic disturbances on electrical systems at the earth's surface," *Advances in Space Research*, vol. 22, no. 1, p. 17–27, 1998.

[22] W. H. Campbell, "Correlation of sunspot numbers with the quantity of s. chapman publications," *Transactions, American Geophysical Union*, vol. 49, no. 4, p. 609, 1968.

[23] A.-u. Raheem, H. Cavus, G. C. Coban, A. C. Kinaci, H. Wang, and J. T. Wang, "An investigation of the causal relationship between sunspot groups and coronal mass ejections by determining source active regions," *Monthly Notices of the Royal Astronomical Society*, vol. 506, no. 2, p. 1916–1926, 2021.

[24] D. B. Dhuri, S. M. Hanasoge, and M. C. Cheung, "Machine learning reveals systematic accumulation of electric current in lead-up to solar flares," *Proceedings of the National Academy of Sciences*, vol. 116, no. 23, p. 11141–11146, 2019.

Optimalization of 64x100 Gb/s DWDM system with DQPSK modulation

[1]*Norbert ZDRAVECKÝ (3rd year)*
Supervisor: [2]Ľuboš OVSENÍK

[1,2]Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

[1]norbert.zdravecky@tuke.sk, [2]lubos.ovsenik@tuke.sk

***Abstract*— This article describe simulation of 64 channel DWDM (Dense Wavelength Division multiplex) system, which has been designed in the OptiSystem program. Optimalization of 64x100 Gb/s DWDM include nonlinear phenomena affecting the transmission. The improvement of EDFA amplification is included in the scheme, which show us better transmission properties of developed system. Proposed optimalization of simulation DWDM and optical amplification is also included in this article. In chapter IV you can find achieved results of BER and Q factor parameters for 100 Gb/s systems with DQPSK (Differential Quadrature Phase-Shift Keying) modulation.**

***Keywords*— BER, DPSK, DQPSK, DWDM, Q-factor.**

I. INTRODUCTION

In the mid-1990s, dense WDM (wavelength division multiplexing - DWDM) systems came with 16 to 40 channels and spacing from 100 to 200 GHz. At the end of the 20th century, DWDM systems reached the point where they were able to transmit 64 to 160 channels in parallel, separated by 50 or even only 25 GHz [1,2]. At the same time as the length density increased, the systems developed their flexibility of configuration, transmission functions and control capabilities. By using wavelength division multiplexing technology, several wavelengths can be simultaneously multiplexed through the fiber [2,3]. Today, systems with 128 and 160 wavelengths are used. Many fiber optic applications, such as long-haul DWDM systems, require optical signal amplification. In the past this meant using expensive electronic amplifiers at midpoints every 100 km. Today's modern long distance DWDM systems use several advanced optical components to replace multiple amplifiers with a single optical device. Once laboratory curiosities, these optical devices, called optical amplifiers, are now seeing widespread use in field deployments. One of the most common optical amplifiers is the Erbium Doped Fiber Amplifier (EDFA).

II. DWDM TECHNOLOGY

In the case of DWDM, the frequency of the light is determined, not the wavelength. Because DWDM is based on long-distance transmission using an optical amplifier, only C-band and L-band are used, which can be easily amplified, and the wavelength is very finely divided to increase the transmission capacity [4,5]. DWDM use serial optical components such as DWDM filters, MUX and DEMUX units, AWG modules, and optical add/drop modules for wavelength division [5].

III. AMPLIFICATION METHODS

Erbium Doped Fiber Amplifier (EDFA) was developed to replace numerous electronic amplifiers with fewer optical amplifiers, reducing overall system cost and complexity [4]. EDFAs also allow easy system upgrades by adding additional sources at different wavelengths and combining them onto a single fiber using a DWDM multiplexer. EDFA consists of a fiber doped with erbium ions with an energy level of $Er^{3+.}$ Ions can absorb radiation at the wavelength of 980 nm and generate radiation in the spectrum around 1550 nm, specifically in the C and L bands [5,6]. This fiber uses a laser pump signal at the level of 980 nm or 1480 nm for pumping. An integral part is a pair of optical isolators, where the isolator at the input of the amplifier prevents the radiation created during spontaneous emission to disturb the amplification [7]. The second isolator at the output prevents the emission of laser beams, and thus possible damage to the amplifier itself, when the beam is reflected into the amplifier [8,9]. A combination of several key factors has made the EDFA amplifier the number one choice for today's optical communication systems based on the WDM. The most important factors are:

- high-performance semiconductor laser pumps,
- polarization independence,
- there is no crosstalk when amplifying WDM signals.

Due to the large EDFA gain of around 30-50 dB, simple construction and reliability, the EDFA amplifier is mainly used as a line amplifier [9,11]. Less used as a booster or preamplifier.

.

IV. SIMULATION OF 64 CHANNEL DQPSK

Wavelength of transmitter is 1550 nm. Spacing frequency is 200 GHz. Power of laser is 2 mW. EDFA (Erbium Doped Fiber Amplifier) amplify gain is 15 dB. Optical SMF(Single mode fiber) is 50 km long and DCF (dispersion compensating fiber) is 10 km long [10]. The Q factor represents the tolerance of the system in dB. We evaluate BER, Q factor and eye height.

Table I shows first channel result of our system. On Fig. 1 is shown scheme of a 64x100 Gb/s system that was simulated in OptiSystem. On Fig. 2 is shown DQPSK subsystem simulation scheme in OptiSystem.

TABLE I RESULTS OF SIMULATION (FIRST CHANNEL) DWDM

Parameters	Channel 1
Q factor	13.2 dB
Minimal BER	3.201e-15
Eye height	2.4256e-6

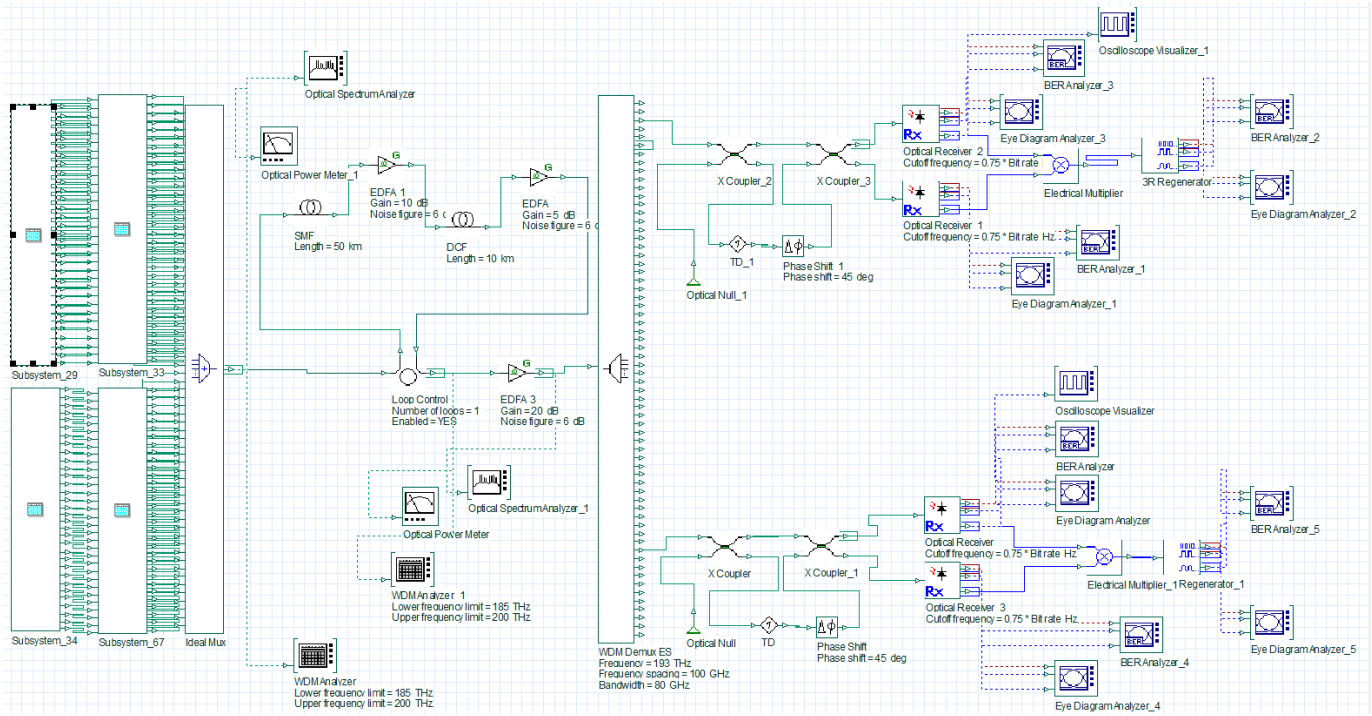


Fig. 1 General diagram of the 64-channel system from the OptiSystem program

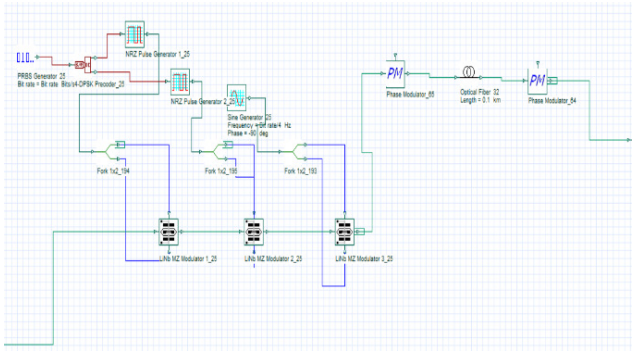


Fig. 2 DQPSK modulation scheme

V. FUTURE WORK

Our future work is defined by dissertation thesis:

1) Optimalization of high-capacity Ultra-DWDM systems for the investigation of nonlinear effects (SPM, XPM and FWM) for transmission speed above 40 Gb/s and 100 Gb/s in fully optical transmission networks.
2) Implementation of modified advanced optical modulation formats to reduce the impact of nonlinear phenomena for Ultra-DWDM optical transmission systems.
3) Optimization of optical network with the optical amplification of transmission paths in high-capacity Ultra-DWDM optical transmission systems.

ACKNOWLEDGMENT

This research was funded by the Slovak Research and Development Agency, research grant no. APVV-17-0208 and VEGA 1/0260/23.

REFERENCES

[1] AGRAWAL, Govind P. Optical Fiber Communications: Optics and Photonics. 6th Edition. Cambridge: Academic Press, 2019. ISBN 9780128170427

[2] GUENTHER, Robert D. Modern Optics Simplified. Oxford: Oxford University Press, 2019. ISBN 978-0-19884-286-6.

[3] DONG, Po. at al. Silicon In-Phase/Quadrature Modulator With On-Chip Optical Equalizer. Journal of Lig-htwave Technology. 2015, 33(6), 1191-1196. ISSN 0733-8724.

[4] HASANUZZMAN R.; M. ISLAM; M. SEKH A. Study of Pumping Methods In Erbium Doped Fiber Amplifier (EDFA). International Journal of Engineering Research & Technology, 2017, vol. 6, no. 8, p. 261-264.

[5] MUBARAKAH N., D. FADHILAH and SUHERMAN, "Point to Point Communication Link Design by Using Optical DWDM Network," 2020 4rd International Conference on Electrical, Telecommunication and Computer Engineering (ELTICOM), 2020, pp. 265-268, doi: 10.1109/ELTICOM50775.2020.9230479

[6] CHANDRA, Aniruddha and Chayanika BOSE. Series solutions for pi/4-DQPSK BER with MRC. International Journal of Electronics. 2011, 99(3), 391-416. Available on: doi:10.1080/00207217

[7] GILL, M. S., S. DEWRA (2017). Analysis of DWDM System Using DPSK Modulation Technique with Raman–EDFA Hybrid Optical Amplifier. Journal of Optical Communications, 38(4). doi:10.1515/joc-2016-0056

[8] SADINOV M., et al, Modelling and performance analysis of DWDM passive optical network, 2021, IOP Conf. Ser.: Mater. Sci. Eng. 1032 012003

[9] HUSZANÍK, Tomáš, Ján TURÁN and Ľuboš OVSENÍK, "Evaluation of CP-DQPSK modulated DWDM system with highly nonlinear fiber in C band," Radioelektronika 2019 : 29th International conference. – Pardubice, p. 304-307. ISBN 978-1-5386-9321-6..

[10] SIMMONS J. M, Optical Network Design and Planning, p.529, 2017. ISSN: 1935-3847, doi: 10.1007/978-3-319-05227-4.

[11] HAN, Sungmin, Jaeseok LEE, Taesoon KWON and Ji-Woong CHOI. Performance analysis on DPSK modulation using symbol repetition and interleaving. Communication systems. 2018, 31(11), 1-16. Available on: doi:10.1002/dac.3589

Modeling and Experimental Identification of Nonlinear Dynamical Systems

[1]*Tomáš TKÁČIK (2[nd] year),*
Supervisor: [2]Ján JADLOVSKÝ, Consultant: [3]Anna JADLOVSKÁ

[1,2,3]Dept. of Cybernetics and Artificial Intelligence, FEEI TU of Košice, Slovak Republic

[1]tomas.tkacik@tuke.sk, [2]jan.jadlovsky@tuke.sk, [3]anna.jadlovska@tuke.sk

***Abstract*—The presented paper deals with the results obtained during the second year of my PhD study in the field of nonlinear dynamical system experimental identification. The methodology of the experimental identification is presented in four consecutive steps with a description. Application of the presented methodology is summarized in approximation model identification of the magnetic levitation system. Also, the ALFRED application is described with the outlook of using it in experimental identification.**

***Keywords*—Nonlinear Dynamical Systems, Experimental Identification, Modeling, Digital Twin, Detector Control System**

I. Introduction

The present trend in the rapid development of computer technologies results in the system design phase being increasingly approached fully digitally using the CAD software. Many software tools allow the created design to be simulated in a virtual environment to test the performance of the design without the need to build a physical model [1]. This opens the question of the credibility of the results that solely rely on the system model. The model is therefore a virtual representation of the real system - digital twin [2].

The model can be obtained in two standard ways - *analytical identification* and *experimental identification* [1]. Analytical identification (or first principles modeling) uses physics laws to derive a model in the form of differential or difference equations. Conversely, experimental identification is based on data. It is the growing number of sensors and the interconnection of systems that enable one to collect a large amount of data to create model by means of experimental identification [3].

II. Previous analysis and achieved results

The experimental identification is a set of data-driven methods and algorithms. It is an iterative approach consisting of four steps depicted in Fig. 1. Each step is supported by a priori knowledge gained from experience with similar systems or as a result of analytical identification [1].

The data collection step combines experiment design, data gathering, and data preprocessing. The experiment design takes into consideration dynamics and constraints of the system to design the most informative experiment that is also safe and economically viable [4]. The data gathering makes use of sensors to digitally record data from the system. The data is later preprocessed using signal processing algorithms to mitigate the effects of the noise or to transform data [3].

The model structure selection step is influenced by the amount of information taken from the analytical identification.

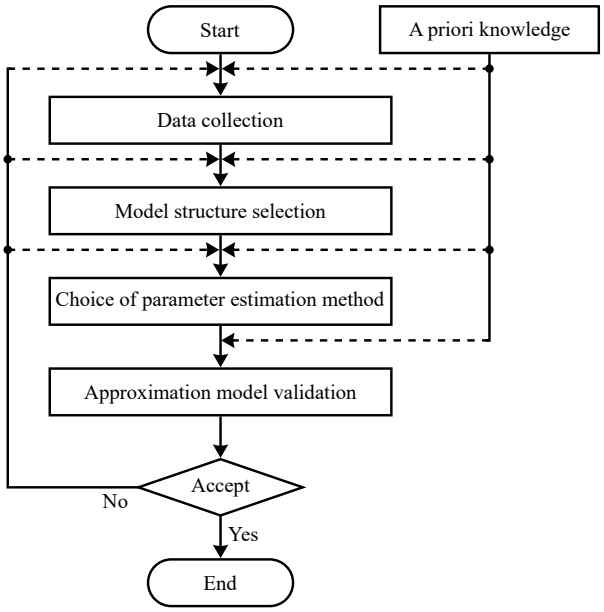


Fig. 1. Experimental identification loop [3].

In case the model structure is obtained using analytical identification, the model is considered to be a gray-box model and only the parameters are to be estimated [5]. On the other hand, black-box models are obtained using data-driven methods to estimate parameters and the model's internal structure [4].

The choice of parameter approximation method step is dependent on the selected model structure. Generally, the parameter value estimation minimizes the output prediction error $\varepsilon(k) = y(k) - \hat{y}(k)$ [6]. Methods of local or global optimization are utilized to estimate parameter values.

Finally, the goal of model validation is to prove the approximation model meets application requirements (e.g., to design a controller or perform diagnostics). Qualitative methods are strictly goal-oriented and results are drawn from an expert's judgment [4]. Quantitative methods on the other hand use statistical metrics and are better suited for the performance comparison of multiple approximation models [3].

III. Solved tasks and results

During the second year of the PhD study, the author has focused on the identification of the magnetic levitation system (MLS) created by the Humusoft company. The mathematical model has been derived first using the physics laws. The resulting mathematical model included parameters that could not

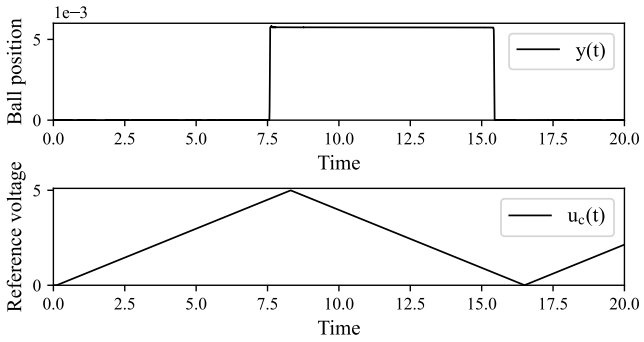


Fig. 2. Sample of triangle signal used to identify static properties of the magnetic levitation system.

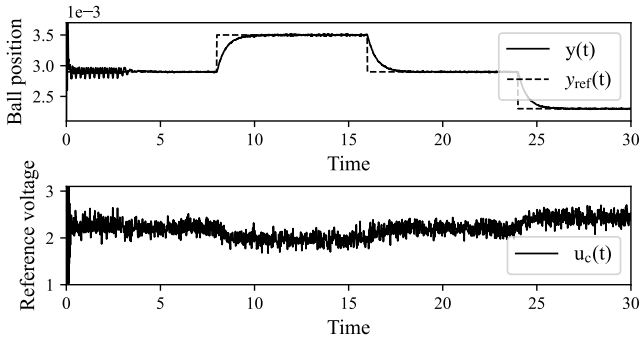


Fig. 3. Application of the LQI controller on the real magnetic levitation system.

be measured directly. Therefore it was necessary to perform experimental identification to be able to design stabilizing controller for this system. Further details will be published in the Tkáčik et.al. article in the AEI journal that is currently under the review.

Experimental identification of the MLS was performed according to the methodology presented in the previous chapter. The dynamics of the MLS stands out as very fast and unstable. Since it is not possible to create even a crude controller for this system, we opted for the experimental identification in the open loop setup. Triangle and pseudorandom binary (PSRB) signals were applied to measure the static and dynamic properties of the system respectively [7]. Sample of the triangle signal applied to the MLS is shown in Fig. 2.

The analytically identified model structure was chosen as the final structure of the approximation model. Initial parameter values were estimated from static analysis data. Later these values were refined using the nonlinear least squares method in conjunction with PSRB data [8].

The approximation model validation was performed on the different realization of the PSBR signal. The qualitative approach of visual comparison between approximate and real output was inconclusive. For the final validation, it was decided to design a stabilizing LQI control algorithm. Application results of the designed LQI control algorithm evaluated on the real system are shown in Fig. 3. The created approximation model in conjunction with designed communication interfaces is the digital twin of the MLS.

The presented MLS is part of the research and development platform at the CMCT&II that the author has been finalizing during the last year. This platform will be used for the validation of intelligent identification and control methods.

The author is also actively participating in the project called *Experiment ALICE on LHC in CERN: Study of strongly interacting matter at extreme energy densities* in collaboration with the CERN. The author is working on the FRED program module of the ALFRED application that is part of the Detector Control System (DCS) used to control physics experiments. The FRED module is used to generate command sequences to control detector components based on requests from WinCC OA [9].

Last year's contribution was mainly in the performance and testing of the FRED module. This included the creation of scripts for automatic testing and the tool to simulate ALF behavior thus mitigating the need for actual hardware components of the detector during testing.

IV. Future research steps

Future research will be devoted to the exploration of the knowledge transfer from gray-box models to black-box models and to improve the overall quality of the approximation model. In [10] the idea to combine gray-box models with neural network-based black-box models proved to be successful. Therefore, we see potential in the combination of analytical and experimental identification approaches, especially with the use of AI-based methods. Last but not least, the author will be participating in the commissioning of the FRED program module as part of the DCS and also in the experimental identification of detector electronics.

V. Conclusion

The presented paper deals with a methodology for experimental identification. It presents experimental identification as the sequence of four steps supported by a priori knowledge. The experimental identification of the magnetic levitation system as the gray-box model is presented alongside the improvements made to the ALFRED application.

Acknowledgment

This work has been supported by the project ALICE experiment at the CERN LHC: The study of strongly interacting matter under extreme conditions (ALICE KE FEI TU 0410 / 2022).

References

[1] L. Ljung, T. Glad, and A. Hansson, *Modeling and Identification of Dynamic Systems*. Studentlitteratur AB, 2021.

[2] M. Liu, S. Fang, H. Dong, and C. Xu, "Review of digital twin about concepts, technologies, and industrial applications," *Journal of Manufacturing Systems*, vol. 58, pp. 346–361, 2021, digital Twin towards Smart Manufacturing and Industry 4.0.

[3] O. Nelles, *Nonlinear system identification: from classical approaches to neural networks, fuzzy models, and gaussian processes*. Springer Nature, 2020.

[4] J. Schoukens and L. Ljung, "Nonlinear system identification: A user-oriented road map," *IEEE Control Systems Magazine*, vol. 39, no. 6, pp. 28–99, 2019.

[5] L. Ljung, "Perspectives on system identification," *Annual Reviews in Control*, vol. 34, no. 1, pp. 1–12, 2010.

[6] A. Jadlovská, *Modelovanie a riadenie dynamických procesov s využitím neurónových sietí*. Edícia vedeckých spisov FEI TU Košice, 2003.

[7] G. Takács, J. Vachálek, and B. R. Ilkiv, *Identifikácia sústav (System Identification)*. Edícia vysokoškolských učebníc STU Bratislava, 2014.

[8] T. Tkáčik, *Contribution to methods, algorithms, and means of identification of nonlinear dynamical systems*. Thesis for the dissertation examination, TU, 2023.

[9] M. Tkáčik, J. Jadlovský, S. Jadlovská, L. Koska, A. Jadlovská, and M. Donadoni, "Fred—flexible framework for frontend electronics control in alice experiment at cern," *Processes*, vol. 8, no. 5, p. 565, 2020.

[10] T. Rogers, G. Holmes, E. Cross, and K. Worden, *On a Grey Box Modelling Framework for Nonlinear System Identification*. Springer, 03 2017, pp. 167–178.

Assessing the automated cross-dataset classification of Parkinson's disease from voice

[1]*Máté HIREŠ (4th year),*
Supervisor: [2]*Peter DROTÁR*

[1,2]Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

[1]mate.hires@tuke.sk, [2]peter.drotar@tuke.sk

***Abstract*—This study investigates the use of shallow learning for the detection of Parkinsonian dysarthria from speech data. The proposed approach utilizes the XGBoost classifier, while processing near 500 acoustic features. In this study we consider two datasets with different language. The model yielded high accuracies while internally validating, even if the two datasets are merged. However, in case of external validation the model fails to generalize. By adding training samples from the test environment helps to improve the models performance. With five added subjects per class the model was able to improve the classification accuracy by 16% and 18%.**

***Keywords*—machine learning, Parkinson's disease, feature extraction, pathological voice, cross-dataset generalization**

I. Introduction

Parkinson's disease (PD) is the second most common neurodegenerative disorder which affects the brain's neurotransmitter systems, resulting in motor and non-motor impairments [1], [2].

Many PD patients experience speech impairments, particularly in the initial stage of the disease [3]. The collection of speech disorders is referred to as dysarthria, its specific type, the Hypokinetic dysarthria (HD) is a speech disorder affecting individuals with PD. HD is characterized by slow, harsh, and monotone speech, slurred speech, reduced volume and pitch, and difficulty with pronunciation and articulation [4], [5].

Although there is no cure for HD, speech therapy can help individuals manage their symptoms and improve communication skills. Deep brain stimulation and other treatments can also be used in conjunction with speech therapy [6].

Diagnosing PD can be a long and costly process requiring specialist evaluations. The high cost of diagnosis can result in some patients being unable to receive a proper medical evaluation, putting their health at risk.

The use of acoustic voice analysis for Parkinson's disease detection is a promising method as it offers important biomarkers for PD diagnosis in the initial stage of the disease [7]. Various techniques have been applied successfully in PD detection, however, most of them are only internally validated [8], [9]. Therefore, they are not suitable in real-world applications, since that would require testing the approach on different data from other environments.

II. Data

A. PC-GITA dataset

The PC-GITA dataset comprises of recordings from 50 individuals with PD and 50 healthy subjects with 25 male and 25 female individuals in each group. Five different speech tasks were recorded at a 44.1-kHz frequency. The detailed description of the speech tasks is presented in [10].

B. Italian Parkinson's voice and speech dataset

The Italian Parkinson's voice and speech (ItalianPVS) database contains recordings from 50 individuals, 22 healthy controls and 28 subjects with PD. One recording session contains 4 tasks, which were recorded at a 16-kHz frequency. The description of the recordings can be found in [11].

In this study we only consider one task from both datasets: the recordings of the sustained phonations of the vowel /a/. In the PC-GITA dataset one subject has three recordings of the given vowel, in the ItalianPVS there are two repetitions.

III. Methods and reults

A. Data processing and feature extraction

To unify the two datasets, the speech signals were downsampled to 16 kHz. Approximately 500 acoustic features were extracted from the input data, considering also the deltas, mean and standard deviation of the segment-level features. We consider the following features, which were selected based on the recent literature [12], [13], [14], [15], [16]: *formants (F1, F2, F3, and F4), shimmers, jitters, linear spectral frequencies, linear spectral coefficients (LPC), detrended fluctuation analysis (DFA) features, pitch period entropy (PPE), harmonics-to-noise ratios (HNRs), log-energy values, zero crossing rate (ZCR), energy, entropy of energy, mel frequency cepstral coefficient (MFCC), spectral centroid, spectral spread, spectral entropy, spectral flux, spectral roll-off, chroma vector, chroma deviation, log mel-spectrogram, Morlet continuous wavelet transform (CWT), F0 contour, and intensity.*

Features were then scaled to the range [0, 1].

B. Experimental setup and results

The first experiment was to train and test the model in the same environment, thus we divided the dataset into train and test set, while we ensure that recordings of one patient are either in the training or the testing set. We considered the connected dataset as well, while data from each dataset were presented also in the training and testing set.

To further examine the models ability to generalize, in the second experiment we trained the model using one dataset, while tested it on the other.

We determined to use the XGBoost classifier for all of our experiments to validate the proposed approach, mainly because of its success on different machine learning challenges.

We tuned the hyperparameters using the grid search cross-validation process in order to achieve the best performance.

To counteract the potential bias resulting from random test data selection, we employed 10-fold cross-validation.

As it can be seen in Table III, the approach was successful in case of the first experiment. The model was able to detect PD accurately even in the connected dataset, reaching 85.47% classification accuracy.

Train	Test	ACC	SEN	SPE
PC-GITA	PC-GITA	87.67	88.67	86.67
ItalianPVS	ItalianPVS	94.00	94.67	93.50
Mixed*	Mixed*	85.47	86.86	84.08

TABLE I
RESULTS OF FIRST EXPERIMENT: INTERNAL VALIDATION
*CONTAINS DATA FROM BOTH DATASETS

The results of the cross-dataset validation are presented in Table II. They signalize, that the model fails to detect PD in a new environment. Thus, in the next experiments we aimed to teach the model some characteristics of the new environment by moving additional training data from the target environment to the train set. We evaluated several experiments by showing the model data from one to five subjects per class (HC and PD) from the target environment. Data from these subjects are excluded from the testing set to avoid bias.

Train	Test	ACC	SEN	SPE
PC-GITA	ItalianPVS	46.77	39.64	55.68
ItalianPVS	PC-GITA	45.20	62.73	27.67

TABLE II
RESULTS OF THE CROSS-DATASET CLASSIFICATION

The achieved results presented in Table III denote, that with additional data from the target environment the model is able to distinguish between PD and HC subject more accurately.

Train	+ test subjects (HC-PD)	Test	ACC(%)	SEN(%)	SPE(%)
PC-GITA	0-0	ItalianPVS	46.77	39.64	55.68
PC-GITA	1-1	ItalianPVS	48.11	72.08	17.86
PC-GITA	3-3	ItalianPVS	53.04	76.00	22.65
PC-GITA	5-5	ItalianPVS	64.02	70.61	55.52
ItalianPVS	0-0	PC-GITA	45.20	62.73	27.67
ItalianPVS	1-1	PC-GITA	50.44	98.87	2.04
ItalianPVS	3-3	PC-GITA	53.33	74.29	35.11
ItalianPVS	5-5	PC-GITA	61.11	87.13	3.83

TABLE III
RESULTS OF THE CROSS-DATASET CLASSIFICATION WITH ADDITIONAL TRAINING DATA MOVED FROM THE TEST SET

We can observe, that by increasing the number of subjects in the training set by moving them from the test set we were able to improve the models generalizability.

IV. DISCUSSION AND CONCLUSIONS

The results presented in this paper indicate that the ML classifier is able to learn the characteristics of PD voices while validating in the same environment. However, when testing with data from an unseen environment, it performs poorly. This can be due to the data being recorded in different environments with different microphone and sampling frequency.

The results also show that adding more training samples from the test dataset helps to improve the models performance. By adding samples from five HC and PD-affected subjects from the target environment the model was able to increase its accuracy from 46.77% to 64.02% and from 45.20% to 61.11%.

The approach presented in this paper has its limitations, one of them is the need of additional data from the target environment. It may be impossible to reach people diagnosed with PD, since doctors can not give out data because of GDPR. A future work may analyze the features characterizing the PD affected voice independently on the recording environment. Another possible solution could be the use of general adversarial networks to generate more diverse data. A bigger multilingual dataset could also be useful for this task.

ACKNOWLEDGMENT

This work was supported by the Slovak Research and Development Agency under contract No. APVV-16-0211.

REFERENCES

[1] M. De Rijk, L. Launer, K. Berger, M. Breteler, J. Dartigues *et al.*, "Prevalence of parkinson's disease in europe: A collaborative study of population-based cohorts. neurologic diseases in the elderly research group." *Neurology*, vol. 54, no. 11 Suppl 5, pp. S21–3, 2000.

[2] M. Trail, C. Fox, L. O. Ramig, S. Sapir, J. Howard, and E. C. Lai, "Speech treatment for parkinson's disease," *NeuroRehabilitation*, vol. 20, no. 3, pp. 205–221, 2005.

[3] L. Hartelius and P. Svensson, "Speech and swallowing symptoms associated with parkinson's disease and multiple sclerosis: a survey," *Folia phoniatrica et logopaedica*, vol. 46, no. 1, pp. 9–17, 1994.

[4] J. B. King, L. O. Ramig, J. H. Lemke, and Y. Horii, "Parkinson's disease: longitudinal changes in acoustic parameters of phonation," *NCVS Status Prog Rep*, vol. 4, pp. 135–149, 1993.

[5] H. L. Klawans, "Individual manifestations of parkinson's disease after ten or more years of levodopa," *Movement disorders: official journal of the Movement Disorder Society*, vol. 1, no. 3, pp. 187–192, 1986.

[6] L. O. Ramig, C. Fox, and S. Sapir, "Speech treatment for parkinson's disease," *Expert review of neurotherapeutics*, vol. 8, no. 2, pp. 297–309, 2008.

[7] B. T. Harel, M. S. Cannizzaro, H. Cohen, N. Reilly, and P. J. Snyder, "Acoustic characteristics of parkinsonian speech: a potential biomarker of early disease progression and treatment," *Journal of Neurolinguistics*, vol. 17, no. 6, pp. 439–453, 2004.

[8] M. Hireš, M. Gazda, P. Drotár, N. D. Pah, M. A. Motin, and D. K. Kumar, "Convolutional neural network ensemble for parkinson's disease detection from voice recordings," *Computers in biology and medicine*, vol. 141, p. 105021, 2022.

[9] M. Hireš, M. Gazda, L. Vavrek, and P. Drotár, "Voice-specific augmentations for parkinson's disease detection using deep convolutional neural network," in *2022 IEEE 20th Jubilee World Symposium on Applied Machine Intelligence and Informatics (SAMI)*. IEEE, 2022, pp. 000 213–000 218.

[10] J. R. Orozco-Arroyave, J. D. Arias-Londoño, J. F. Vargas-Bonilla, M. C. Gonzalez-Rativa, and E. Nöth, "New spanish speech corpus database for the analysis of people suffering from parkinson's disease." in *LREC*, 2014, pp. 342–347.

[11] G. Dimauro, V. Di Nicola, V. Bevilacqua, D. Caivano, and F. Girardi, "Assessment of speech intelligibility in Parkinson's disease using a speech-to-text system," *IEEE Access*, vol. 5, pp. 22 199–22 208, 2017.

[12] J. Rusz, T. Tykalova, M. Novotny, D. Zogala, E. Ruzicka, and P. Dusek, "Automated speech analysis in early untreated parkinson's disease: Relation to gender and dopaminergic transporter imaging," *European Journal of Neurology*, vol. 29, no. 1, pp. 81–90, 2022.

[13] M. F. Anjum, S. Dasgupta, R. Mudumbai, A. Singh, J. F. Cavanagh, and N. S. Narayanan, "Linear predictive coding distinguishes spectral eeg features of parkinson's disease," *Parkinsonism & related disorders*, vol. 79, pp. 79–85, 2020.

[14] F. Amato, L. Borzì, G. Olmo, and J. R. Orozco-Arroyave, "An algorithm for parkinson's disease speech classification based on isolated words analysis," *Health Information Science and Systems*, vol. 9, no. 1, pp. 1–15, 2021.

[15] L. Zhang, Y. Qu, B. Jin, L. Jing, Z. Gao, Z. Liang *et al.*, "An intelligent mobile-enabled system for diagnosing parkinson disease: development and validation of a speech impairment detection system," *JMIR Medical Informatics*, vol. 8, no. 9, p. e18689, 2020.

[16] J. C. Vásquez-Correa, J. Orozco-Arroyave, T. Bocklet, and E. Nöth, "Towards an automatic evaluation of the dysarthria level of patients with parkinson's disease," *Journal of communication disorders*, vol. 76, pp. 21–36, 2018.

Exploring Capsule Neural Networks and the Importance of Explainability

[1]*Dominik VRANAY(2nd year),*
Supervisor: [2]*Peter SINČÁK*

[1,2]Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1]dominik.vranay@tuke.sk, [2]peter.sincak@tuke.sk

***Abstract*—To address the limitations of traditional neural networks, capsule neural networks were proposed, which represent objects in a structured, hierarchical way, allowing for better interpretability and capturing complex relationships. Capsule networks model an object as a set of capsules containing information about different aspects such as position, orientation, size, and texture. However, capsule networks face challenges in training, and optimising the network due to non-linear transformations, and difficulty in understanding information propagation. The motivation for capsule networks is to address these challenges and improve interpretability and performance, with the potential to achieve state-of-the-art results in applications such as object recognition and medical image analysis. This paper provides an overview of capsule networks and the current research done in the last year as well as planned future works on this topic as well as the topic of XAI.**

***Keywords*—Capsule Network, explainability, image classification, reconstruction, segmentation**

I. Introduction

Deep neural networks have made significant progress in a wide range of applications, such as image recognition, natural language processing, and autonomous vehicles. However, traditional neural networks have limitations in their ability to model hierarchical relationships between features and often lack interpretability. To address these limitations, capsule neural networks were proposed as a new approach to deep learning [1].

Capsule networks are designed to represent objects and features in a hierarchical and structured way, allowing them to capture complex relationships and making them more interpretable. Capsule networks model an object as a set of capsules that contain information about different aspects of the object, such as its position, orientation, size, and texture. These capsules can be thought of as "object parts" that can be combined to form a complete object [2].

Despite their potential advantages, capsule networks face a number of challenges. One of the key challenges is the difficulty in training and optimizing the network due to the non-linear transformation of capsules, which requires iterative routing procedures to determine the most relevant capsules. Additionally, understanding how information is propagated through the network and how capsules are combined to form a final output can be difficult [3].

The motivation for capsule neural networks is to address these challenges and improve the interpretability and performance of deep neural networks. By providing a structured and hierarchical representation of objects, capsule networks have the potential to capture complex relationships and improve the interpretability of deep neural networks [4]. In addition, capsule networks have shown promise in achieving state-of-the-art results in various applications, such as object recognition and medical image analysis.

II. Background and Previous Findings

There are many fields in which Capsule Networks were used and are able to solve. The main one is image classification. CapsNets have been successfully applied to image classification tasks, including the MNIST dataset and the more challenging CIFAR-10 and CIFAR-100 datasets [5]. CapsNets have been shown to outperform traditional convolutional neural networks (CNNs) on these tasks, particularly when it comes to handling variations in the orientation and position of objects within images [6].

There were many problems with CapsNets which were already solved. The main one would be the high use of memory during training [7]. Another part of this problem was the complexity of the original routing algorithm, which was high and was done in multiple iterations [2]. This made learning slow and computationally very expensive.

This was fixed by creating a simpler routing algorithm between capsule layers, which in the end increased the efficiency of the model and decreased the needed computational power [8]. As these routing algorithms are simpler, the time per epoch decreased as well.

The additional problem, which was partially fixed was the low performance of these models on more complex datasets. As the capsules are using reconstruction, it becomes hard to create a good representation of the input image, which is varied and the needed dimensionality to catch this increases exponentially [9]. This was partially fixed by creating a larger convolutional backbone, which would compress the data and reduce the number of needed capsules [10].

The first capsules also had problems with segmentation tasks on difficult medical data from CT and MRI scans [11]. This was slowly improved by the introduction of a U-Net-like structure created using capsules as well as a custom convolutional routing algorithm, which enabled a very powerful model to be created [12].

III. Research Highlights

The main focus of the research work in the last year was on researching and implementing different routing algorithms,

which were proposed in the last few years. This was done to decide the most successful one, which would be used in future research. Self-Attention routing algorithm and the basic concepts of the Efficient-CapsNet were taken as the base for future research [8].

In the first paper, which is under review process, a modified capsule network SE-CapsNet was proposed, which is based on Efficient-CapsNet with segmentation masks as a reconstruction target. By using segmentation as a target for reconstructions, the models are able to better generalize the information needed for their creation, which in turn helps the performance. In comparison, generalizing the information of fully detailed input images is much harder and many times not possible for our models.

By analyzing multiple models on two different datasets, we showcased that using the segmentation masks improves the performance of the models in classification and they can outperform other capsule and convolutional models. The BCE loss for reconstruction was shown to be the better of the two losses with improvements even on original image reconstructions. The best SE-CapsNet model reached a mean accuracy of 93% on the Oxford pets dataset [13], which is an increase of almost 3% from other capsule architecture and 1.5% from traditional CNNs. A similar increase of 0.5% to the final F1 score performance of 65.5% was recorded on the SIIM-ACR PNEUMOTHORAX dataset [14].

The Oxford pets dataset was chosen as a good representation of real-world images of the outside world with the classes being cats and dogs. The Pneumothorax dataset was used as it shows not so varied X-ray images of lungs, which need to be diagnosed. The contrast between these datasets shows the general improvement of this technique.

The use of segmentation was then further explored by envisioning an explanation module, which would help with the interpretability of the model and help increase the trust index of the doctors, which would use this model. This is possible by analysing the results of the segmentation and the classification together and checking the highlighted parts of the image, which should contain the details needed from the model to give the result. The model can be adjusted to have 0 false negatives, which will enable the doctors to use it as a screening tool. For the final diagnosis, they would only have a fraction of the images and each would have its own segmentation, in which the model would find something so the doctor can only focus on the needed region. A number of created reconstructions can be seen in Fig. 1.

Another research was on the possibility of using the original reconstructions created by capsule models to explain parts of the networks. This was done by creating a small capsule model trained on a simple dataset Fashion MNIST [15]. The resulting capsules were then slowly deformed by separating one feature of the reconstructed capsule and chaining it in small intervals. The resulting reconstructions were analysed and each feature was given an explanation of what it changes in the image.

This research into the deformations of the capsule dimensions was presented at the SAMI 2023 conference and won the Baltazár Frankovič Young Researcher Award.

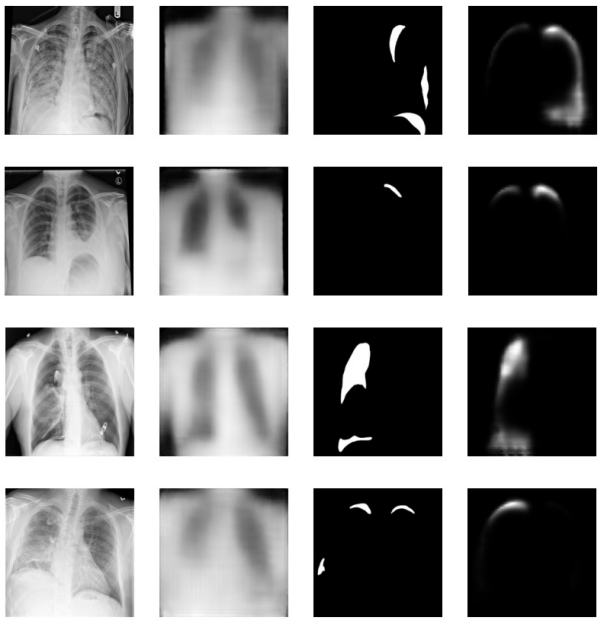


Fig. 1. Examples of the reconstructions using capsule models with different targets. The first column shows the original input for all models, the second its reconstruction created by a model with input as a target, the third shows the segmentation mask of the input image, and the last shows the reconstruction of the mask by a different model.

IV. Planned Work

The first key component of the planned work is focused on analyzing the parse trees within capsule neural networks. Parse trees provide a structured representation of the input data and the processing steps performed by the neural network. By analyzing the parse trees, we aim to gain a deeper understanding of how information is propagated through the network and how different features are combined and transformed as they move through the network layers.

To achieve this goal, we will be using a combination of visualization techniques and statistical analysis to identify patterns and relationships within the parse trees. They will also be developing new tools and methods for interpreting and visualizing parse trees, which will help to uncover insights and anomalies in the network's behaviour [16].

The second key component of the planned work is focused on improving the routing algorithm used in capsule neural networks. The routing algorithm is responsible for determining how information is passed between different capsules in the network, and it plays a critical role in the performance of the network.

To improve the routing algorithm, we will be creating a separate feature for the existence of the object. This new feature will allow the routing algorithm to better account for the presence or absence of an object in the input data, which is expected to improve the network's performance in classification tasks and other related applications [17].

The third and final key component of the planned work is focused on creating a library of routing algorithm implementations. This library will provide researchers and developers with a range of routing algorithm options that can be easily integrated into their own capsule neural network projects.

The library will be designed to be flexible and extensible, allowing users to easily swap out different routing algorithms and compare their performance on different datasets. We will also be providing documentation and examples to help users get started with the library and understand how to use it effectively.

REFERENCES

[1] G. E. Hinton, A. Krizhevsky, and S. D. Wang, "Transforming auto-encoders," in *Int. Conf. on artificial neural networks*. Springer, 2011, pp. 44–51.

[2] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," *Advances in neural information processing systems*, vol. 30, 2017.

[3] M. Mitterreiter, M. Koch, J. Giesen, and S. Laue, "Why capsule neural networks do not scale: Challenging the dynamic parse-tree assumption," *arXiv preprint arXiv:2301.01583*, 2023.

[4] N. Xie, G. Ras, M. Van Gerven, and D. Doran, "Explainable deep learning: A field guide for the uninitiated," *arXiv e-prints*, pp. arXiv–2004, 2020.

[5] D. Vranay, "Capsule neural networks - future of deep learning?" in *22nd Scientific Conference of Young Researchers*, 2022, pp. 170–173.

[6] F. D. S. Ribeiro, G. Leontidis, and S. Kollias, "Capsule routing via variational bayes," in *Proc. of the AAAI Conf. on Artificial Intelligence*, vol. 34, no. 04, 2020, pp. 3749–3756.

[7] G. E. Hinton, S. Sabour, and N. Frosst, "Matrix capsules with em routing," in *Int. Conf. on learning representations*, 2018.

[8] V. Mazzia, F. Salvetti, and M. Chiaberge, "Efficient-capsnet: Capsule network with self-attention routing," *Scientific Reports*, vol. 11, no. 1, pp. 1–13, 2021.

[9] Y. Wang, L. Huang, S. Jiang, Y. Wang, J. Zou, H. Fu, and S. Yang, "Capsule networks showed excellent performance in the classification of herg blockers/nonblockers," *Frontiers in pharmacology*, vol. 10, p. 1631, 2020.

[10] I. J. Jacob, "Capsule network based biometric recognition system," *Journal of Artificial Intelligence*, vol. 1, no. 02, pp. 83–94, 2019.

[11] R. LaLonde and U. Bagci, "Capsules for object segmentation," *arXiv preprint arXiv:1804.04241*, 2018.

[12] M. Tran, L. Ly, B.-S. Hua, and N. Le, "Ss-3dcapsnet: Self-supervised 3d capsule networks for medical segmentation on less labeled data," in *2022 IEEE 19th Int. Symposium on Biomedical Imaging (ISBI)*. IEEE, 2022, pp. 1–5.

[13] O. M. Parkhi, A. Vedaldi, A. Zisserman, and C. V. Jawahar, "Cats and dogs," in *IEEE Conf. on Computer Vision and Pattern Recognition*, 2012.

[14] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *Proceedings of the IEEE Conf. on computer vision and pattern recognition*, 2017, pp. 2097–2106.

[15] H. Xiao, K. Rasul, and R. Vollgraf. (2017) Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms.

[16] D. Erhan, Y. Bengio, A. Courville, and P. Vincent, "Visualizing higher-layer features of a deep network," *University of Montreal*, vol. 1341, no. 3, p. 1, 2009.

[17] C. Molnar, *Interpretable machine learning*. Lulu. com, 2020.

Evaluation of the possibility of connecting a RES to the distribution power system

[1]*Jozef HUMENÍK (3rd year)*
Supervisor: [2]*Jaroslav DŽMURA*

[1,2]Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic

[1]jozef.humenik@student.tuke.sk, [2]jaroslav.dzmura@tuke.sk

***Abstract* - The current direction of the development of the field of energy, smart grid and microgrid is also closely related to the development of Renewable Energy Sources (RES). From the European regulations to the national level, it is possible to perceive massive support for the use of RES through various projects. It is an effort to connect more and more such resources. This issue can be perceived from different points of view, from the point of view of an official in the European union, through the departments/ministries of individual countries to the point of view of the Distribution System Operators (DSO). This article is mainly devoted to the point of view of the Distribution System Operator on the given issue from a technical point of view. The specific topic is the determination, or determination of the limit/maximum power of the renewable energy source that should be connected to an electric power system. To ensure a safe, optimal and constant supply of electricity to final consumers without unnecessary undesirable phenomena such as non-compliance with quality parameters of electricity or, in the worst case, interruption of electricity supply (deterioration of SAIDI - System Average Interruption Duration Index and SAIFI - System Average Interruption Frequency Index indicators).**

***Keywords* - RES, balance point, short-circuit power, calculation of power.**

I. INTRODUCTION

The massive increase in the number of especially small (local) sources and thus the increase in installed power can be observed all over Slovakia, from small households to business entities. Currently, almost everyone who has enough funds can arrange for the installation of a source. Of course, the connection conditions of the distribution system operator must be met. From a business point of view, this is an excellent business model, e.g., for electrical installation companies that install and connect sources at customers. For a comprehensive assessment of the issue, it is necessary to consider the appropriateness of connecting such several sources and the impact on the electrical network, that is, the overall impact on the system.

II. CONNECTING RESOURCES TO THE DISTRIBUTION SYSTEM

Distribution System Operators have defined technical conditions for connecting resources to the distribution power system and establish criteria, or allowable values of feedback effects of sources on the distribution power system. From the analysis of the technical conditions of connection of the Slovak Distribution System Operators (western, central, and eastern Slovakia), the requirements for parallel cooperation of sources are identical mainly in the following criteria:

- permissible voltage changes in steady state,
- permissible voltage changes when switching the source,
- flicker,
- harmonic currents.

Each producer of electricity is obliged to comply with the conditions and criteria for the connection of its sources to the Distribution System (DS), at the same time, each source must meet the specified conditions for voltage changes in the steady state (2%) and during switching (3%). [1][2][3][4][5][6][7][8]

A. The procedure for determining the limit/maximum power of RES

The mathematical determination of the limit/maximum power of RES intended for connection to the distribution system consists of three levels:

- calculation of the maximum power of the source with respect to the outlet of the med-voltage(MV) transformer,
- calculation of the maximum connectable power of the resources with respect to the free connectable the capacity in the 22 kV system determined regarding the power of High voltage (HV)/Extra-High voltage (EHV) transformers,
- calculation of the maximum connectable power of the resources with respect to the free connectable capacity in the 110 kV system determined regarding the power of the Ultra-High voltage (UHV)/Very-High voltage (VHV) transformers. [1][2][3][4][5][6][7][8]

Calculation of the maximum power of the source with respect to the outlet of the MV transformer

During the calculation, the assessment of the source is also considered with respect to the permitted voltage change in the steady state and during switching. It is not possible to consider harmonic currents in the calculation due to the production of different harmonic currents with different inverters. Fig. 1. shows the algorithm for calculating the maximum power of the source with respect to the outlet of the transformer.

The calculation should be made for all transformer outlets, and the sum of the powers thus determined on all 22 kV outlets represents the calculated limit power with respect to the HV (MV) outlets.

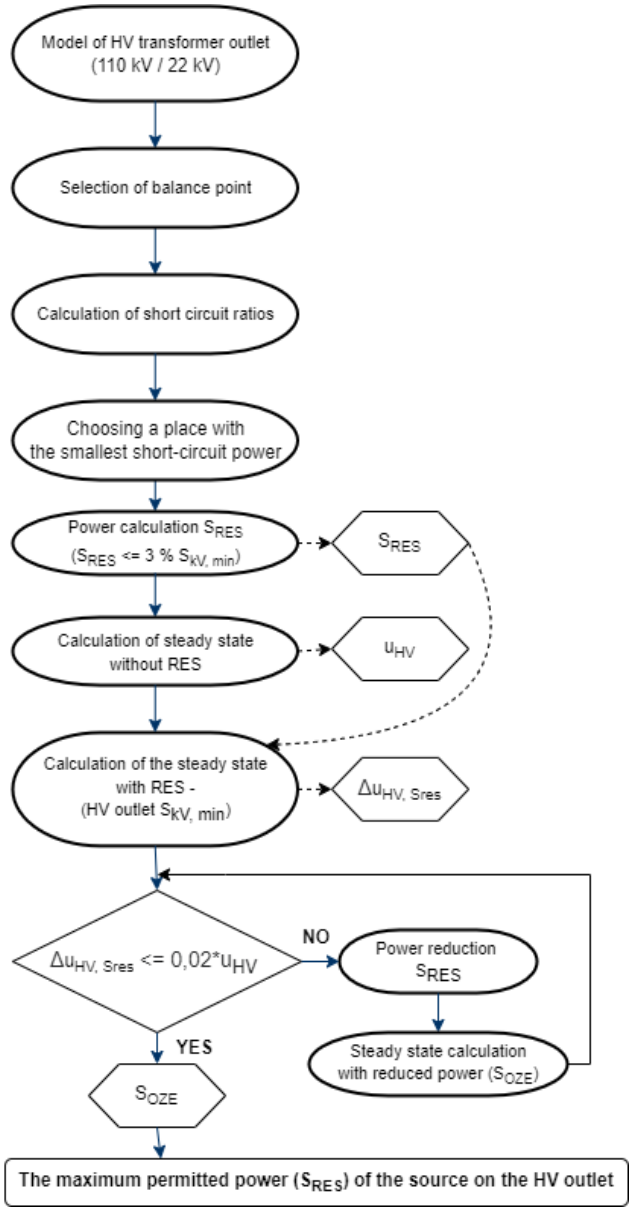


Fig. 1. Algorithm for calculating the maximum permissible power of the source

Calculation of the maximum connectable power of the resources with respect to the free connectable the capacity in the 22 kV system determined regarding the power of HV/EHV transformers

The maximum power of the sources regarding the power of the 110 kV/22 kV transformers is determined according to the following relationship:

$$P_{\max 22kV/outlet} = \left(\sum P_{i(n-1)} \cdot k_{Tr} + P_{consumers}\right) \cdot k_E \quad (1)$$

$P_{i(n-1)}$ - is the sum of the installed outlets of 110 kV/22 kV transformers in the assessed area excluding the transformer with the highest output (criterion n-1),

k_{TR} - coefficient considering the optimal load of the transformer, connect ability of renewable sources of electricity to the electricity system

$P_{consumers}$ - total minimum consumption of the assessed area,

k_E - reduction coefficient considering small scattered production. This coefficient enables the creation of a certain power reserve for sources with low installed power (up to 0.10 MW). [1][2][3][4][5][6][7][8]

The determination of the maximum power of the sources regarding the free connectable capacity in the 110 kV system is determined regarding the power of the MV/VHV transformers

To determine the maximum power of the sources regarding the power of the 400 kV/110 kV (or 220 kV/110 kV) transformers of the given 110 kV area, the following relationship is used:

$$P_{\max,110kV} = \left(\sum P_{i(n-1)} \cdot k_{TR} + P_{consumers}\right) \cdot k_E \quad (2)$$

$P_{i(n-1)}$ - the sum of the installed powers of transformers 400kV/110kV, or 220kV/110kV in the assessed area excluding the transformer with the highest power (criterion n-1). [1][2][3][4][5][6][7][8]

III. Conclusion

Based on various analyses, investigations, and mathematical calculations, it was confirmed that the issue should be considered from a comprehensive point of view. From the point of view of Distribution System Operators, source connection requirements and local distribution influences, it is possible to connect such a few small sources that would exceed the capabilities of the currently built system itself. The procedure for determining the limit/maximum output of RES, which should be connected to the electricity system, considers local impacts, but also overall (systemic) impacts. The obtained results of the limit/maximum power calculation prove that the largest volume of installed power can be connected in terms of local effects at the level of individual outlets of the 22 kV network. Analysing towards the transmission system with respect to the free capacities of the distribution transformers, a decreasing tendency of performance can be observed. The total amount of installed power that can be connected to the electricity system is limited mainly by system influences. The limit of the installed power of RES in the electricity system is correctly determined if the ability of the system is mainly assessed from the point of view of system effects.

References

[1] MEDVEĎ, D.: *Modelovanie prechodných dejov pri pripojovaní rozptýlených zdrojov energie v prostredí EMTP ATP (angl. Modeling of transient events when connecting scattered energy sources in an EMTP ATP environment)*. In: Elektroenergetika, Vol. 3, No. 7, 2010, p. 15-18. ISSN 1337-6756

[2] SZATHMÁRY, P.: *Kvalita elektrickej energie (angl. The quality of electrical power)*. Vol. 1. Banská Bystrica: PRO, s.r.o., 2003. p. 122. ISBN 80-89057-04-7

[3] Technical conditions of connection of the individual Distribution System Operators (DSO)

[4] HUMENÍK, J.: *Author's own notes from studies/conferences/workshops and work/practical skills*. 2022.

[5] Act No. 309/2009 Coll. on the support of renewable energy sources and high-efficiency combined production

[6] Act No. 251/2012 Coll. on Energy and on change of certain acts

[7] Standard PNE 33 3430-0 and PNE 33 3430-2

[8] Standard STN EN 50160: 2011

A comprehensive overview of the study and analysis of Agile methodologies

[1]*Emira Mustafa Moamer ALZEYANI (2nd year)*
Supervisor: [2]Csaba SZABÓ

[1,2]Dept. of Electrical Engineering and Informatics, FEI TU of Košice, Slovak Republic

[1]emira.mustafa.moamer.alzeyani@tuke.sk, [2]csaba.szabo@tuke.sk

***Abstract*— The world of software project management is very vast. It has many different areas that build on each other to work towards challenging any developer's difficulties on the job. Like any field of life, this field is always subject to constant evolution in its working methods, style and technology that depends on it. Here we will discuss our studies on the Agile methodology in several ways. First, to help us build a model directed to a specific category of developers who do not have high experts in the field of project development, or we can say that they are in the early stages of the world of software development, like university students. And because it is an important stage in developing themselves in software development; they must learn and master everything new in project management to work effectively and efficiently in the future.**

***Keywords*— Agile methodology, Artificial Intelligence, Datasets.**

I. INTRODUCTION

Agile methodology is the key to envisioning a profitable and sustainable future for any organisation or company, whatever its size or market. Agile methodologies can also be seen as a humane business focusing on customer value and increasing team productivity and collaboration. In addition, agile values and principles will help you achieve this culture of collaboration.

Agile is not just a process or methodology for running a business. Instead, it is a mindset that embraces challenges and views failure as learning opportunities and game-changers; which teaches us the value of working and thinking in an Agile methodology; It also focuses on adaptable and energetic leaders in their leadership style. This is where leaders learn to lead themselves first and create an environment that allows Agile to thrive.

The emergence of Agile as a broad global movement extends far beyond software development projects. However, it has been discovered that the only way for organisations to deal with today's turbulent client-driven market is to switch to an Agile methodology. Agile enables organisations to master continuous change in project management. It allows businesses to grow exponentially in an increasingly volatile, uncertain, complex and uncertain world.

It makes you avoid financial risks, spend months or years, and realise small mistakes that could lead to big problems. It depends on building trust between employees, the employee's self-confidence, his efforts and his work focusing on dealing directly with the goals that serve customers and providing quick and gradual solutions.

II. OUR PRELIMINARY RESULTS

At this stage, we decided to shed light on the studies we carried out and published them in several journals. These studies helped us build a preliminary design of the needs we must consider in the next model design stage.

A. Study on the most commonly used methodologies

In this research [1], [2], we identified the latest techniques most used in project management under the Agile methodology, highlighted the most important characteristics and clarified the differences and the basis for selection. The focus was on the methodologies of Scrum, SAFe, and DAD. We concluded that the Scrum model is one of the most smooth methodologies. It has high efficiency in the work of projects with a narrow scope and the difficulty of working in the case of a large team distributed geographically. The SAFe methodology is one of the most complex frameworks in the way of working. It has many strict rules to be dealt with while developing a specific product. It is very structured, which decreases flexibility and agility. DAD is one of the minor complex frameworks to work with; it has more than one framework in the DAD framework that can be worked on according to the system to be implemented or according to the team. The team can work on more than one project at the same time, and each project has a particular framework in line with the requirements of the product.

B. The importance of building a work team in project management

This study [3] Presents the importance of teamwork and the responsibility it engages in during work, the methods of communication between them and relying on each other.

And how teamwork works through self-management to speed up work. Due to the difficulty of working alone on developing or implementing a specific product, we must work on developing ourselves in teamwork to get the desired result: delivering the product to the required place within the specified time and budget.

C. Study a group of students in a college

The study [4] is about a group of students from the faculties of Informatics in Tripoli, Libya. The aim was to identify difficulties during project development and the methodology used. The study was conducted with a questionnaire directed

to a specific group of students in the study's last years. Fifty results were collected. The results obtained were based mainly on the Waterfall methodology.

The reasons for choosing the Waterfall methodology were the small team, the clarity of the requirements, and the lack of continuous change.

The conclusion from this study is the importance of depicting the method of training developers on new and developed methods in project management. To make developers work according to the most effective methodologies in software development through training in environments that require a large team and the development of large and more complex projects, which helps them determine how to choose the appropriate model for the project to be developed.

And through the paper [5], we have explained the importance of the planning stage and the possibility of using the AMEISE system and highlighted the reason for using it in teaching software project management. This tool allows small groups to plan for each simulation and then let them interact with the simulated events as they simulate to achieve as many micro-objectives as possible, such as budget, length, and other requirements.

D. Study on open-source datasets

This paper has been accepted by Acta Electrotechnica et Informatica journal but has not been published yet. It revolves around studying and analysing open-source datasets to help us prove the actual extent of the Agile methodology in the various circumstances that any programmer can encounter. The open-source datasets that worked under the Agile methodology, especially the Scrum model, were analysed using Python with the Exploratory Data Analysis model (EDA).

The results demonstrated the actual methodology by delivering the project with the required requirements and the project's success with all measures—the results obtained from the Tscrum_peter_Balaz datasets [6] and the TAWOS [7].

We have reached the point where we can overcome any difficulty through good management, work control and interdependence among the team. Thus, we guarantee the effectiveness and success of the Agile methodology.

III. CONCLUSION

We conclude from these studies with several main factors we can rely on to build a model based on the Agile methodology. First, it includes good coordination between team members, user participation in the work at every stage, continuous communication between team members, and good coordination between the project manager and the team homogeneously. Also, factors are developing the team to work under artificial intelligence. And how we can take advantage of it; it helps us accelerate work performance. And the most important thing is that we can apply it and use it in training for the educational stage to help students get used to high-level work during the study period.

IV. FUTURE WORK

After the previous studies, the points and scope of the next work were clear, which were clarified in the following questions and hypotheses

1. How can we design a model that is easy to handle and teach based on the above features?

AI could be used as a supervisor or support for management. Thus, we think a hypothesis on the possibility of including an artificial supervisor in agile processes at the level of control will be evaluated as truth.

The second part of our hypothesis addresses the usage of deep neural networks as the best option for agile project management supervisor implementation. We will experimentally prove this hypothesis by applying EDA to find the best option for a simulation core element of the generic model presented in [8]. An open question remains whether the developers' personnel skills should also be considered. And if yes, what weight should be assigned to them?

2. What is the best stage we can use AI in project management?

We will focus on planning and workflow tracking since these are the most related stages to the expected result of our first hypothesis. We assume that support based on track is the most advantageous feature for all team members. Such support might be the most beneficial to teaching and training (as it decreases supervisor workload). An open question remains whether it is possible to divide the stages of any agile methodology in the above way. If our hypothesis on the possibility of such divisions fails, all stages must be considered when constructing the model.

3. Can we rely on the developed Agile methodology to build a project based on AI?

We assume the developed extended agile methodology will be reliable enough for application. This hypothesis will be proved in-vitro as a specific agile project planning assignment in the frame of the Software Project Management teaching the subject at DCI FEI. We will apply statistical methods for evaluation, similar to the study presented in [4]. We might also need to extend the validation to other universities if the significance of personnel skills, habits and culture is strong while examining question 1.

REFERENCES

[1] E. M. M. Alzeyani and . C. Szabó, "Comparison of Agile Software Project Management Methods," in *22nd Scientific Conference of Young Researchers*, 2022.

[2] E. Alzeyani and . C. Szabo, "Simulation Model for Agile Software Project Management," in *2021 19th International Conference on Emerging eLearning Technologies and Applications (ICETA)*, 2021.

[3] E. Alzeyani, "DIGITAL SKILLS IN AGILE TEAMWORK ENVIRONMENT," in *Skills for (post) Digital Society, Košice (Slovakia),* 2022.

[4] E. M. Alzeyani and C. Szabó, "A study on the methodology of Software Project Management used by students whether they are using an Agile or Waterfall methodology," in ICETA 2022, pp.22-27, 2022.

[5] C. Szabó, B. Osif and E. M. M. Alzeyani, "Project Planning Support for Waterfall Software Project Management Simulations," in *XIII International Conference of Information Technology and Development of Education ITRO 2022*, Zrenjanin, 2022.

[6] P. Baláž, E. Mustafa Moamer Alzeyani and C. Szabó, "Web Application for Simulation of Agile Software Projects – Stage One: Team Builder," *Journal of Applied Technical and Educational Sciences,* vol. 12, pp. 1-11, 2022.

[7] V. Tawosi, A. Al-Subaihin, R. Moussa and F. Sarro, "A Versatile Dataset of Agile Open Source Software Projects," 2022.

[8] I. J. De Silva, S. Rayadurgam and M. P. E. Heimdahl, "A reference model for simulating Agile processes," 2015.

Exploring the Diagnosis of Parkinson's Disease through Offline Handwriting: An Artificial Intelligence Approach

[1]*Matej Gazda(4^{th} year),*
Supervisor: Ján Plavka

[1,2]Dept. of Mathematics and Theoretical Informatics, FEI TU of Košice, Slovak Republic

[1]matej.gazda@tuke.sk, [2]jan.plavka@tuke.sk

***Abstract*—Parkinson's disease (PD) is a prevalent neurodegenerative condition with a significant global impact. In this study, we evaluate the feasibility of using artificial intelligence-powered models for diagnosing PD through offline handwriting analysis. Our findings are compared with previous literature to provide a comprehensive overview of the field's current state. The results highlight the challenges posed by limited data and low heterogeneity in developing a usable model for clinical applications. Our analysis reveals that the model lacks generalizability when train and test datasets are acquired from the same site using similar equipment. This study highlights the need for robust data sources and diverse data samples to improve the accuracy of PD diagnosis through handwriting analysis.**

***Keywords*—Parkinson's disease diagnosis, offline handwriting, online handwriting, deep learning**

I. Introduction

Diagnostic decision support systems use various data input forms to help health professionals make more informed and objective decisions. Data might come in the form of medical images, such as x-ray or magnetic resonance, temporal data, such as electroencephalography and electrocardiography, or even just normal images representing people, such as images of their writing or their posture.

The accuracy of Parkinson's disease (PD) diagnosis by experts, as reported in the scientific literature, varies from 75% to 92% and is dependent on the expertise of the examiner [1], [2]. A pathological study revealed that a maximum of 25% of patients diagnosed with PD had been misdiagnosed, even in the advanced stages of the disease. The symptoms of PD result in a substantial decrease in the quality of life for affected individuals and also result in a significant economic impact on society, with annual estimated costs of 52 million dollars in the United States alone [3].

Handwriting is a learned motor and cognitive function of the brain, and as such, it can serve as an indicator of cognitive or motor dysfunction. It has been proven that handwriting has been established as an early indicator of PD.

The current state-of-the-art for objective evaluation of PD using computerized methods can be divided into online and offline handwriting approaches. Online handwriting captures signals from a digitizing tablet to describe the handwriting process [4], [5]. Although this method has proven to be successful, it requires an expensive tablet and a specialized pen. The trend in research is moving towards offline handwriting evaluation, where only a pen and a smartphone, which is easily accessible, are required. Deep learning techniques have demonstrated promising results in PD diagnosis from offline handwriting. In this study, we demonstrate the current state-of-the-art results for PD diagnosis by deep learning, compare it to our approaches, and finally show that studies need further validation to be proven clinically usable.

II. Literature Review

The diagnosis of PD is a vivid research area in both medical and computer science domains. Interestingly, PD can be diagnosed in various ways, investigating either handwriting or speech impariaments, while some doctors are also inclined to include medical imaging tools to reject the possibility of other diseases with similar symptoms.

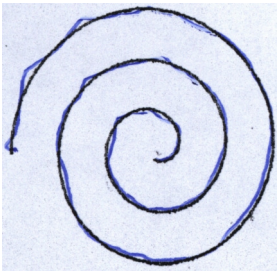
Diaz et al. [6] proposed a sequence-based dynamic handwriting analysis by levearing temporal data with one-dimensional convolutions and BiGRUs. Diaz et al. [7] proposed a model based on a static representation with embedded dynamic information. Among the features that might support PD diagnosis are speed, velocity, in-air or on-surface movement, jerk, pressure, azimuth, etc. Both approaches required spatiotemporal data that had to be acquired with specialized pens. Moetesum et al. [8] proposed a model leveraging features extracted from sole images by convolutional neural networks (CNNs), and those features were later fed into a support vector machines classifier. Our works [11], [10] leveraged novel multiple-fine-tuning methods with advanced ensembling methods to detect PD disease purely from handwriting images that could be taken from a cellphone. Galaz et al. [12] acquired a new dataset of PD and HC (WorldPaHaW) from various sites in the United States, Hungary, Spain, and Czech that consisted of language-independent spiral and language-dependent sentences. The study also compared models based on carefully handcrafted features and CNNs from static images.

III. Results

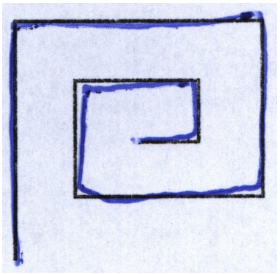
Alongside multilingual WorldPaHaW [12], two datasets have been constructed for handwriting analysis: NewHandPD [13] and PaHaW [4]. The NewHandPD dataset comprises two tasks: drawing meanders and creating Archimedean spirals (see Fig. 1a and 1b). In the case of the PaHaw dataset, the handwriting is recorded as x and

Task	Diaz [6]	Diaz [7]	Moetesum [8]	Pereira [9]	Gazda [10]	Gazda [11]	Galaz [12]
Spiral (language leave-one-out)	–	–	–	–	–	–	BACC - (38% – 71%)
Spiral (NewHandPD)	94.44	–	–	76.26	92.7	**96.3**	–
Meander (NewHandPD)	91.11	–	–	80.75	94.7	**94.4**	–
Spiral (PaHaW)	**93.75**	75	76	–	85.8	88.5	–
l (PaHaW)	**96.25**	64.16	62	–	68	71.25	–
le (PaHaW)	**88.75**	58.33	57	–	74.7	78.8	–
les (PaHaW)	**90**	71.67	60	–	72.7	72.5	–
lektorka (PaHaW)	**93.75**	75.41	60	–	76.1	81	–
porovnat (PaHaW)	**91.25**	63.75	51	–	76	77.26	–
nepopadnout (PaHaW)	**92.5**	70	68	–	78.5	91.88	–

TABLE I: Overview of deep learning performance for offline PD diagnosis. Standard accuracy is reported, however for Galaz [12], balanced accuracy is reported since the dataset is not balanced



(a) NewHandPD Spiral



(b) NewHandPD Meander

y coordinates, which are utilized to generate images. The PaHaW dataset is an aggregation of eight distinct handwriting tasks, including the creation of Archimedean spirals, letters, syllables, and words. Little work has been done on PD diagnosis with models trained on multilingual datasets. Galaz's WorldPaHaW was acquired from 143 PD patients and 151 healthy controls from the Czech Republic, the United States, Colombia, and Hungary. The dataset consisted of language-independent tasks (spiral) and sentence-writing tasks.

In Table I, we report state-of-the-art results of PD diagnosis from both online and offline handwriting. The table shows that for PaHaW tasks, the model for online handwriting by Diaz et al. [6] outperformed other models on PaHaW tasks. Our work in [10], [11] improved previous work significantly and established state-of-the-art results for PD diagnosis purely from static images (offline handwriting). Interestingly, all those studies had been trained and tested on datasets that didn't have enough heterogeneity. A study by Galaz [12] showed that models working with handcrafted features or features extracted from static images yielded a performance on spiral tasks with balanced accuracy ranging from 38% to 71%, depending on the external validation. A study performed leave-out-language-out validation, which means that the model was, for instance, trained on data from the United States, Czech Republic, and Colombia and was tested on data from Hungary. The study by Galaz showed the importance of possessing external datasets acquired from different sites.

IV. Future work

Presently, graph neural networks have been on the rise in the research domain since they can have rotational and translational invariance, which is an essential feature for handwriting analysis. Those properties are essential because the image scan can be captured from different angles, with different cameras and different positions of the letters on the paper. Furthermore, one might also include handcrafted features that could be embedded in the graph nodes. Graph Attention Layers [14] can be leveraged to find one graph-level representation of the handwritten task and diagnose PD.

References

[1] A. Schrag, Y. Ben-Shlomo, and N. Quinn, "How valid is the clinical diagnosis of parkinson's disease in the community?" *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 73, no. 5, pp. 529–534, 2002.

[2] J. Jankovic, "Parkinson's disease: clinical features and diagnosis," *Journal of neurology, neurosurgery & psychiatry*, vol. 79, no. 4, pp. 368–376, 2008.

[3] W. Yang, J. L. Hamilton, C. Kopil, J. C. Beck, C. M. Tanner, R. L. Albin, E. Ray Dorsey, N. Dahodwala, I. Cintina, P. Hogan *et al.*, "Current and projected future economic burden of parkinson's disease in the us," *npj Parkinson's Disease*, vol. 6, no. 1, p. 15, 2020.

[4] P. Drotar, J. Gazda, and Z. Smekal, "An experimental comparison of feature selection methods on two-class biomedical datasets," *Computers in Biology and Medicine*, vol. 66, pp. 1 – 10, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0010482515002917

[5] P. Drotar, J. Mekyska, I. Rektorova, L. Masarova, Z. Smekal, and M. Faundez-Zanuy, "Decision support framework for parkinsons disease based on novel handwriting markers," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 23, no. 3, pp. 508–516, May 2015.

[6] M. Diaz, M. Moetesum, I. Siddiqi, and G. Vessio, "Sequence-based dynamic handwriting analysis for parkinson's disease detection with one-dimensional convolutions and bigrus," *Expert Systems with Applications*, vol. 168, p. 114405, 2021.

[7] M. Diaz, M. A. Ferrer, D. Impedovo, G. Pirlo, and G. Vessio, "Dynamically enhanced static handwriting representation for parkinson's disease detection," *Pattern Recognition Letters*, vol. 128, pp. 204–210, 2019.

[8] M. Moetesum, I. Siddiqi, N. Vincent, and F. Cloppet, "Assessing visual attributes of handwriting for prediction of neurological disorders—a case study on parkinson's disease," *Pattern Recognition Letters*, vol. 121, pp. 19–27, 2019.

[9] C. R. Pereira, D. R. Pereira, G. H. Rosa, V. H. Albuquerque, S. A. Weber, C. Hook, and J. P. Papa, "Handwritten dynamics assessment through convolutional neural networks: An application to parkinson's disease identification," *Artificial intelligence in medicine*, vol. 87, pp. 67–77, 2018.

[10] M. Gazda, M. Hireš, and P. Drotár, "Multiple-fine-tuned convolutional neural networks for parkinson's disease diagnosis from offline handwriting," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 52, no. 1, pp. 78–89, 2021.

[11] M. Gazda, M. Hires, and P. Drotár, "Ensemble of convolutional neural networks for parkinson's disease diagnosis from offline handwriting," *IGS2021: The 20th Conference of the International Graphonomics Society*.

[12] Z. Galaz, P. Drotar, J. Mekyska, M. Gazda, J. Mucha, V. Zvončák, Z. Smekal, M. Faundez-Zanuy, R. Castrillon, J. R. Orozco-Arroyave *et al.*, "Comparison of cnn-learned vs. handcrafted features for detection of parkinson's disease dysgraphia in a multilingual dataset," *Frontiers in Neuroinformatics*, p. 35, 2022.

[13] C. R. Pereira, S. A. T. Weber, C. Hook, G. H. Rosa, and J. P. Papa, "Deep learning-aided parkinson's disease diagnosis from handwritten dynamics," in *Proceedings of the SIBGRAPI 2016 - Conference on Graphics, Patterns and Images*, 2016, pp. 340–346.

[14] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," *arXiv preprint arXiv:1710.10903*, 2017.

Collecting and preprocessing ECG data from microcontrollers for future classification

[1]*Dávid Vaľko (2nd year)*
Supervisor: [2]Norbert Ádám

[1,2]Dept. of Computers and informatics, FEI TU of Košice, Slovak Republic

[1]david.valko@tuke.sk, [2]norbert.adam@tuke.sk

***Abstract* — This paper describes ECG signal acquisition from microcontrollers and further signal preprocessing. Cardiovascular illnesses are the number one cause for most deaths worldwide. Myocardial infarction and cerebrovascular diseases are the leading causes of cardiovascular mortality. Despite certain risk factors, healthy people can have a myocardial infarction or brain stroke as well. Medical staff can spot early warning symptoms in ECG recordings. Early detection allows treatment for patient before suffering serious health problems. This paper describes a strategy to obtain relevant ECG recording for future classification, thus avoiding health problems. Creating an affordable home diagnostics machine gives the opportunity to have patient's heart checked regularly.**

***Keywords*— computer science in medicine, cardiovascular disorders, ECG, signal preprocessing**

I. Introduction

The electrocardiogram (ECG) is a cutting-edge method that measures the electrical activity of the human heart. Because it is a noninvasive measurement, it can be performed by anyone without putting the patient at risk of being harmed [1]. This indicates that the ECG diagnosis is risk-free and does not cause any discomfort.

The ECG presents an examination that checks the heart from a particular angle and detects the electrical activity in millivolts. A 12-lead ECG is the most frequent type of electrocardiogram recording that is employed in medical practice. It is true that in order to conduct a full medical examination on a human being, it is necessary to get a 12-lead ECG. Despite this fact, many life-threatening diseases can be recognized even with a single-lead ECG [2].

Our successes were already published on SAMI 2022 and SAMI 2023 conferences. In 2022, we have published proper configuration of convolutional neural network in terms of ECG classification [3]. In 2023, we have proposed ways of acquiring favorable ECG signal for later classification [4].

II. Limb and chest leads

Based on acquired results from the past [3-4], we want to further explore the possibilities of acquiring and classifying ECG recordings, that could help in a patient's diagnostics, thus preventing serious health damage. The notion of Einthoven's Triangle, which is seen in Fig. 1, can be used to provide an explanation for the concept of ECG in general.

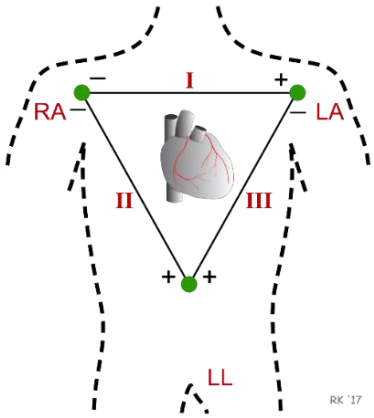


Fig. 1 The concept of Einthoven's triangle (source: https://www.cvphysiology.com/Arrhythmias/A013a)

Thanks to this theory, the spots for placing ECG electrodes have been standardized worldwide. This triangle guarantees that the electrical conductivity of a heart will be examined from every angle in the frontal plane [5].

There are several different denotations included in the figure, namely LA, RA, and LL. These acronyms refer to the left arm, the right arm, and the left leg, respectively. The electrical difference that exists between the two electrodes [6] is to determine a single lead on an ECG. Because of this knowledge, the bipolar leads (I, II, and III) are determined by applying the following formulas:

- *Lead 1 (I) = LA − RA*
- *Lead 2 (II) = LL − RA*
- *Lead 3 (III) = LL – LA*

By using leads I, II, and III, we are able to determine what are known as augmented leads. These are referred to as the aVR, aVL, and aVF notations respectively. The letter "a" represents the augmented format, whereas the letter "V" refers to the vector form.

The fact that these leads are obtained through the use of computing means that they do not offer any sort of new information; rather, they provide a different perspective on leads that already exist. The following equations are used to determine the number of augmented leads:

- $\mathrm{aVL} = \frac{Lead\ I - Lead\ III}{2}$
- $\mathrm{aVF} = \frac{Lead\ II + Lead\ III}{2}$
- $\mathrm{aVR} = \frac{(-1)\ (Lead\ I + Lead\ II)}{2}$

Recordings of an ECG taken using a single-lead device, such as a smartwatch or a microcontroller, are likely to be highly disturbed due to the specific circumstances involved [7] such as patient movement or powerline interference (PLI). The measurement of electrical activity between the LA and RA (positive and negative) electrodes is the foundation of a single-lead electrocardiogram. By moving these electrodes on different parts of body, we might discover or compute new leads.

III. Using AD8232 for ECG recording

The AD8232 device may be linked to any controller directly, including Arduino and Raspberry Pi. The AD8232 is a commonly used sensor for ECG recordings in terms of developing low-cost devices. There have been many studies that have utilized this sensor for various applications [8]. By using AD8232 it is possible to acquire data in digital form and save them in a PC for later use [9]. It is critical to acquire an ECG signal of the highest possible quality. Despite the fact that we have access to filtering techniques and digital denoisification, it is best to eliminate signal disturbances as early as possible in the process. The signal of poor quality is usually labeled as "unreadable" and may not be used in classification.

Based on the known facts, we were able to propose specific lead placement (Fig. 2) using AD8232 in order to obtain Leads I, II and III. Our experiment proved that it is possible to obtain different leads just by adjusting individual leads.

This also lead us to believe, that it is also possible to obtain chest leads V1 – V6 in similar way.

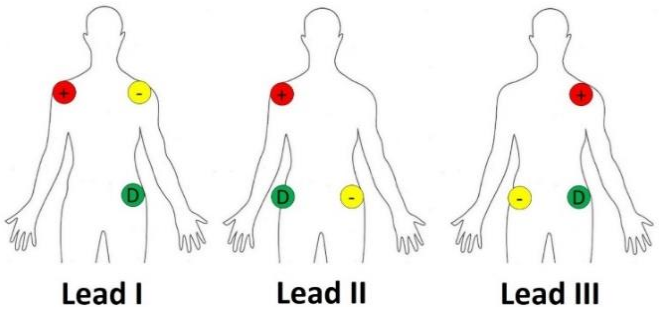


Fig. 2 Proposed limb leads placement using AD8232

Using measurements taken with an AD8232 device, we were able to get the raw ECG data from the limb leads. The Fourier transformation was executed on each limb lead during this stage, and the results demonstrated that the signal contains unwanted noise mostly at a frequency of 50 Hz. The experiment was repeated on the remaining leads and the results were the same. To address this issue, we have used raw ECG signal in .csv format and imported the values in python language (Fig. 3). Later, we used filtering techniques to remove frequencies ranging from 48-52 Hz to eliminate PLI. We have used python library *matplotlib.pyplot* to visualize acquired ECG data. Then we have used SciPy library and created a bandstop filter using function *signal.butter()*.

While continuing our research, we have also discovered that the ECG recording can be further polished by removing all frequencies above 40 Hz. An important thing to note here is that no information was lost or removed in recording. Every part of ECG was maintained and wasn't distorted.

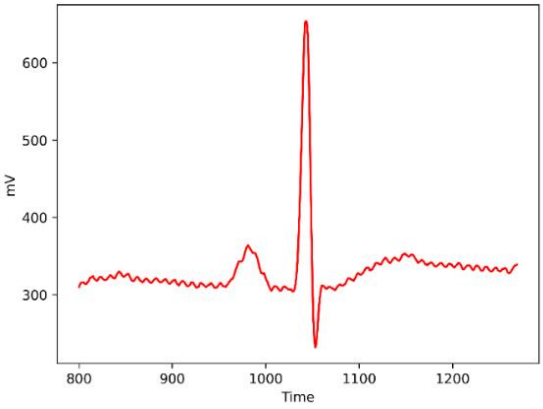


Fig. 3 Raw ECG signal with only one QRS

This procedure lead us to high-quality ECG recordings, which are clear and diagnosable. Such recordings (Fig. 4) can be fed to neural networks for classification in the future. Overall, our approach offers a more comprehensive and accurate understanding of the proposed topic.

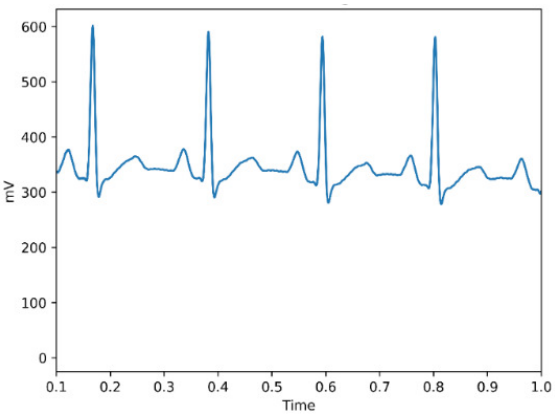


Fig. 4 Filtered ECG signal

IV. Future experiments

For future experiments, we would like to acquire more ECG recordings from different kinds of devices. We would like to compare these devices and see, which gives ECG signal in best quality. After that, we would explore the possibilities of acquiring chest leads V1-V6 and augmented leads. Early results of this experiment show that it is possible. We aim to develop a more comprehensive and accurate ECG measurement system that can provide valuable clinical insights and improve patient outcomes.

V. Conclusion

We can conclude that we have obtained a high-quality ECG signal from low-cost devices. With proper procedures and techniques, it is possible to record relevant ECG leads. Proper leads placement in order to acquire limb leads was presented in Section III. We have presented a properly filtered and preprocessed ECG signal for future classification.

References

[1] Dupre, Anthony, Sarah Vincent, and Paul A. Iaizzo. "Basic ECG theory, recordings, and interpretation." *Handbook of cardiac anatomy, physiology, and devices* (2005): 191-201.

[2] Beni, Nargess Heydari, and Ning Jiang. "Heartbeat detection from single-lead ECG contaminated with simulated EMG at different intensity levels: A comparative study." *Biomedical Signal Processing and Control* 83 (2023): 104612.

[3] N. Ádám, D. Vaľko and M. Havrilla, "Using Neural Networks for ECG Classification," 2022 IEEE 20th Jubilee World Symposium on Applied Machine Intelligence and Informatics (SAMI), Poprad, Slovakia, 2022, pp. 000367-000372, doi: 10.1109/SAMI54271.2022.9780670.

[4] N. Ádám, D. Vaľko and B. Madoš, "Acquiring favorable ECG signal from low-cost devices," 2023 IEEE 21th Jubilee World Symposium on Applied Machine Intelligence and Informatics (SAMI), Herľany, Slovakia 2023.

[5] Dower, Gordon E., H. Bastos Machado, and J. A. Osborne. "On deriving the electrocardiogram from vectorcardiographic leads." *Clinical cardiology* 3.2 (1980): 87-95.

[6] Zhao, Pei Jun. "Einthoven's Triangle Revisited: A Mathematical Proof." *arXiv preprint arXiv:2205.06772* (2022).

[7] Shi, Fengjun. "A review of noise removal techniques in ECG signals." *2022 IEEE Conference on Telecommunications, Optics and Computer Science (TOCS)*. IEEE, 2022.

[8] Kanani, Pratik, and Mamta Padole. "Recognizing real time ECG anomalies using Arduino, AD8232 and Java." *Advances in Computing and Data Sciences: Second International Conference, ICACDS 2018, Dehradun, India, April 20-21, 2018, Revised Selected Papers, Part I 2*. Springer Singapore, 2018.

[9] Mendes Junior, José Jair Alves, et al. "AD8232 to Biopotentials Sensors: Open Source Project and Benchmark." *Electronics* 12.4 (2023): 833

Design of an RSSI parameter prediction algorithm for hard switching methods in hybrid FSO/RF systems

[1]*Maroš LAPČÁK (3rd year)*
Supervisor: [2]Ľuboš Ovseník

[1,2]Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

[1]maros.lapcak@tuke.sk, [2]lubos.ovsenik@tuke.sk

***Abstract*— This paper deals with the prediction of RSSI (Received Signal Strength Indicator) parameter which is required for hard switching methods in hybrid FSO/RF systems. To ensure almost 100% system availability, it is necessary to predict this parameter correctly. It is affected by several atmospheric influences such as fog, rain, airborne particle concentration, haze, etc. It is these atmospheric influences that have degrading effects on the RSSI parameter. Therefore, it is necessary to investigate and record these influences and correlate them to find out which have the greatest influence on the RSSI parameter. With proper prediction, we can determine well in advance which link will be communicating at that moment. The FSO line is used as the primary one for communication because of its data rate from one side to the other. Thus, the RF line serves only as a secondary (backup) line. For the prediction of RSSI parameter, machine learning methods of Decision Trees along with AdaBoost Regressor was used.**

***Keywords*— decision tree, FSO, hybrid FSO/RF, machine learning.**

I. Introduction

Hybrid FSO/RF systems are a suitable solution for seamless communication. By combining them, the communication gets better fluency and less error rate. The main reason for the emergence of this integrated hybrid FSO/RF system is to improve link availability. The hybrid FSO/RF system offers this improvement by finding a compromise between the reliability of the RF link and the high data rate of the FSO link. The operation process of hybrid FSO/RF systems can be seen in Fig. 1 [1], [2], [3].

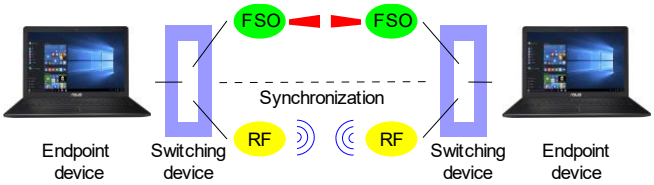


Fig. 1: Principle of operation of a hybrid FSO/RF system.

A. Hard Switching

Hard switching between FSO and RF works by having the transmitter and receiver choose an RF or FSO channel for data transmission. If a situation arises where the primary FSO link is unable to provide data transmission with predefined quality of service (QoS), then the RF link is subsequently used to maintain high reliability. Short-term changes in the environment such as atmospheric turbulence cause sensitivity for switched hybrid FSO/RF systems [4], [5]. This fact influences the mechanism, which consequently requires frequent switching between the two channels depending on the changes in channel conditions, leading to a reduction in system throughput. Based on this, the switching between the optical and RF link is a major issue in the switching system. The selection of the correct channel for data transmission is determined by the coordinated feedback at the transmitter and receiver [6], [7].

B. Decision Trees algorithm

Decision trees algorithm is one of the oldest and most widely used algorithms. It is most often used for classification, but its variant can also be used for regression analysis. A decision tree consists of nodes, with the base node called the root node. The nodes branch further to other nodes to form the structure of the tree. Each node represents a decision according to one selected feature of the classified object. The selected features must make the objects as distinct from each other as possible to be classified as accurately as possible at the end of the tree. Information entropy and the so-called information gain are used for the correct attribute selection, which must divide the data into two branches as different as possible [8], [9], [10].

C. Decision Trees with AdaBoost Regressor

To achieve better results of RSSI parameter prediction, it was desirable to improve the already used Decision Tree algorithm. The improvement of this algorithm could be achieved with the help of the extension of the mentioned Decision Tree regressor and the AdaBoost regressor. In practice, such an improvement of the algorithm shows better prediction results since the AdaBoost regressor uses iterations in its prediction calculation, which it uses to learn from weaker regression methods. However, it should be noted that better prediction results require a much larger time effort [5], [8], [10].

II. Analysis of the Results of the Work

This chapter includes the results of RSSI parameter prediction using the different algorithms, which are Decision Tree and Decision Tree with AdaBoost regression. The results are demonstrated through the generated graphs. The different algorithms have been tested on different sets of input data. The

algorithms were also tested with different ratios of training and test data. Primarily, the work focused on testing different datasets with a ratio of 70:30. Both algorithms ran for program optimality in a loop that was used to determine the training time of each algorithm. This paper focuses in its computations on training data from every 1 minute, every 2 minutes, and every 5 minutes. The best result had Decision Tree algorithm with AdaBoost Regressor. The success of prediction with Decision Tree algorithm with AdaBoost Regressor from every 1st minute was 87%, from every 2nd minute was 82% and from every 5th minute was 72%. Fig. 2 shows graph of the Decision tree algorithm using the AdaBoost regressor which used all input parameters such as Temperature 1, 2, 3, Particle Concentration, Visibility, Pressure, Humidity, Wind Speed and Wind Voltage.

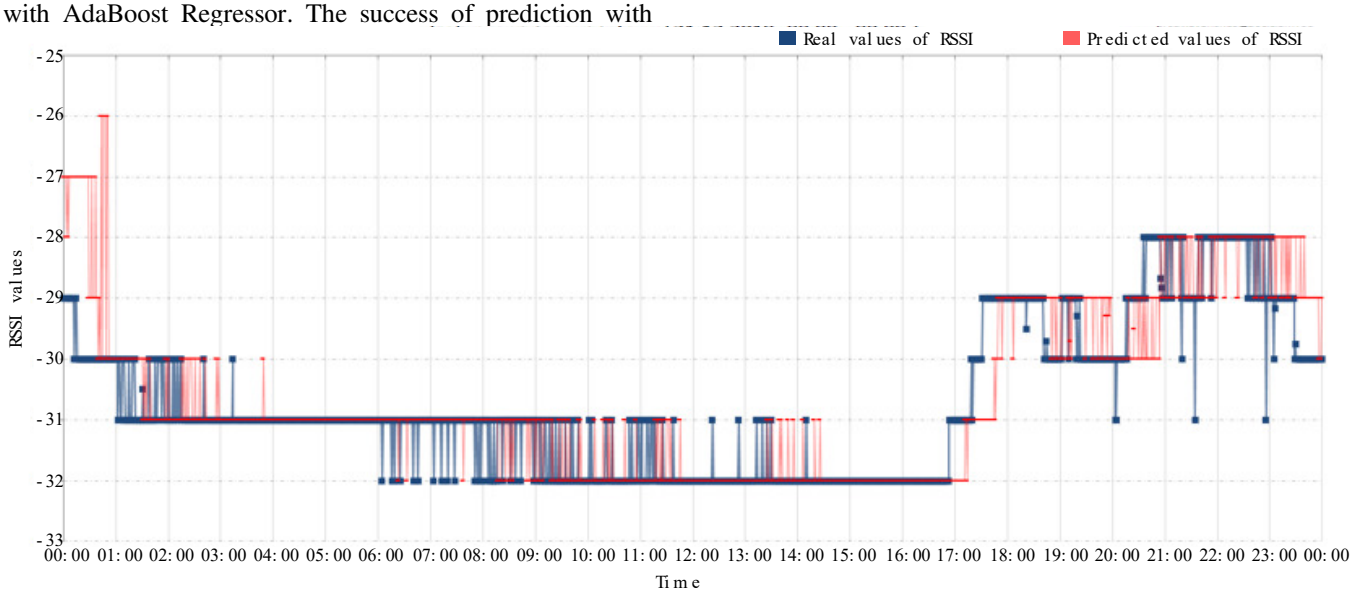


Fig. 2 Comparison of real and predicted RSSI values based on input parameters (from each minute).

III. Conclusion

This paper focused on the prediction of RSSI parameter through two experimental machine learning algorithms, where the set of required input data for RSSI parameter prediction was mainly investigated. The input data analysis required testing different sets of input data as well as testing different ratios of training and test data. Two methods were used namely Decision Tree method and Decision Tree with AdaBoost Regressor. A total of nine different input data sets were tested on both the algorithms, where for each input set, a training set from each of the 1st minute, 2nd minute, and 5th minute was trained. By having many input sets, it was possible to evaluate multiple graphs. The paper demonstrates in the results the graph containing as input dataset all input data. It should be noted the fact that both programs can make a successful prediction only one to two minutes ahead. The example of best results is shown in this paper. The shortcomings of both prediction methods could be eliminated in the future, for example, by calculating the prediction of the individual weather parameters that most influence the RSSI parameter. However, we can say with certainty that based on the obtained results we can correctly predict the RSSI parameter with minimum error and thus ensure smooth operation of the hybrid FSO/RF system based on the hard switching method, since the critical value of switching from FSO to RF link is at the level of -42 dBm.

Acknowledgment

This research was funded by the Slovak Research and Development Agency, research grant no. APVV-17-0208 and VEGA 1/0260/23.

References

[1] Willebrand, H. and Ghuman, W. S., *Free-Space Optics: Enabling Optical Connectivity in Optical Connectivity in Today's Network,*" Indianapolis: Sams Publishing, 2002, 280 p., ISBN: 978-0-672-32248-8

[2] Zihang, Q.; Xiuping, L.; et al., Compact Wideband Circularly Polarized Patch Antenna Array Using Self-Sequential Rotation Technology. *IEEE Antennas and Wireless Propagation Letters,* 2022, Volume 21, No. 4, pp. 700 – 704. DOI: 10.1109/LAWP.2022.3142307.

[3] Arnon, S.; Barry, J.; et al., Advanced Optical Wireless Communication Systems. Cambridge University Press, UK, 2012, 404p.. ISBN: 9780521197878.

[4] Vigneshwaran, S.; Muthumani, I.; Sivanananthа Raja, A. Investigations on Free space optics communication system. *IEEE 2013 International Conference on Information Communication and Embedded Systems (ICICES 2013)*, Chennai, India 2013, pp. 819–824. DOI:10.1109/ICICES.2013.6508300.

[5] Tóth, J.; Ovseník, Ľ.; et al., Long term availability analysis of experimental free space optics system. *2015 International Conference on Systems, Signals and Image Processing (IWSSIP).* 2015, pp. 29 – 32. DOI: 10.1109/IWSSIP.2015.7313929.

[6] Márton, M.; Ovseník, Ľ.; et al., Measurement effect of visibility in experimental FSO system. *2017 IEEE 14th International Scientific Conference on Informatics.* 2017, pp. 249 – 252. DOI: 10.1109/INFORMATICS.2017.8327255.

[7] Song, S.; Liu, Y.; et al., Channel prediction for intelligent FSO transmission system. Optics Express, 2021, Volume 29, No. 17, pp. 27882 – 27899. DOI: 10.1364/OE.433493.

[8] Haluška, R.; Šulaj, P.; Ovseník, Ľ.; et al., Prediction of Received Optical Power for Switching Hybrid FSO/RF System. *Electronics,* 2020 Volume 10, No. 8, 14 p.. DOI: 10.3390/electronics9081261.

[9] Antonios, L..; Konstantinos, P.; et al., Using Machine Learning Algorithms for Accurate Received Optical Power Prediction of an FSO Link over a Maritime Environment. *Photonics*, 2021, Volume 8, No. 6, 17 p.. DOI: 10.3390/photonics8060212.

[10] Tóth, J.; Ovseník, Ľ.; et al., Classification prediction analysis of RSSI parameter in hard switching process for FSO/RF systems. *Measurement*, 2018, Volume 116, pp.602 – 610. DOI: 10.1016/j.measurement.2017.11.044.

Learning turbulent flows with neural network

[1]*Marek RUŽIČKA (4^{th} year),*
Supervisor: [2]*Juraj GAZDA*

[1,2]Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

[1]marek.ruzicka@tuke.sk, [2]juraj.gazda@tuke.sk

***Abstract*—The well established computational fluid dynamics (CFD) software can simulate fluid movement with ease. However, the simulations are time demanding and the software requires expensive hardware, especially graphical units. Recently developed physics informed neural networks (PINNs) are capable to learn specific functions of differential equations and replace the CFD software. In this paper, we show how well established k-ε model can be transformed to a neural network.**

***Keywords*—CFD, PINN, fluids, neural networks**

I. Introduction

Fluid dynamics is the study of the motion of liquids, gases, and plasmas, including their physical properties, behavior, and interactions with other materials. Basically, there are several methods to solve fluid dynamics, including Finite Difference Method (FDM), Finite Element Method (FEM), Finite Volume Method (FVM), and others [1]. In this paper we consider FVM [2]. FVM is a numerical technique used to solve partial differential equations (PDEs) that describe fluid flow, heat transfer, and other physical phenomena in a discrete domain. It involves dividing the domain into a finite number of control volumes, or cells, and discretizing the governing equations by integrating them over each cell. The method then solves the system of algebraic equations resulting from this discretization to obtain a numerical solution. FVM is a conservative method, meaning it ensures that mass, momentum, and energy are conserved across the domain.

In the context of computational methods for solving PDEs, a geometry mesh is a discrete representation of a continuous domain, such as a solid object or a fluid volume. The mesh is typically composed of a finite number of simple shapes, such as triangles, that tile the domain. The use of a mesh allows for the approximation of complex geometries and enables the application of numerical methods to solve equations governing the behavior of fluids, solids, or other materials in the domain.

Fluid dynamics are usually calculated by models built on the Reynolds-Averaged Navier-Stokes (RANS) equations [3]. RANS are a set of time-averaged conservation equations that describe the motion of a fluid. They are derived by decomposing the velocity, pressure, and other properties of the fluid into time-averaged and fluctuating components, and then applying the conservation equations to the time-averaged quantities.

In practice, models built on RANS are used, such as k-ε [4]. The k-ε model is a widely used turbulence model in computational fluid dynamics (CFD) that provides closure for the RANS equations. It assumes that the turbulent kinetic energy (k) and the rate of dissipation of kinetic energy (ε) are the most significant variables in determining the turbulence behavior of a flow field. The k-ε model makes use of transport equations for k and ε that describe how the turbulence energy is transported, dissipated, and produced in a flow field. This model is often used for simulating turbulent flows in industrial applications due to its simplicity and robustness.

Recently, Physics-Informed Neural Networks (PINNs) were proposed [5], which are a class of machine learning algorithms that combine neural networks with physical laws and constraints in order to solve complex physical problems. They are designed to learn the underlying physics of a system by being trained on a small set of observed data and the governing equations that describe the system. PINNs have been applied to a variety of problems in fluid mechanics, heat transfer, and solid mechanics, among others. PINNs can be used to learn the k-ε model by using the governing equations of the model as the physical constraints in the network [6]. The network can be trained to minimize the error between the predicted values of the k-ε model and the actual values.

In this paper, we show the k-ε model and a way to use it in PINN. Then, we compare the results. This research will later be extended to a deep study.

II. From k-ε model to PINN

In k-ε model, k represents kinetic turbulent energy and ϵ is the rate of consumption. k-ε model considers the wind flow to be turbulent, as in the urban environment with multiple obstacles in the wind path. The parameters k and ϵ are determined from the following transport equations:

$$\frac{\partial}{\partial t}(\rho k) + \frac{\partial}{\partial x_i}(\rho k u_i) = \frac{\partial}{\partial x_j}[(\mu + \frac{\mu_t}{\sigma k})\frac{\partial k}{\partial x_j}] + G_k + G_b - \rho\epsilon - Y_M + S_k \quad (1)$$

and

$$\frac{\partial}{\partial t}(\rho \epsilon) + \frac{\partial}{\partial x_i}(\rho \epsilon u_i) = \frac{\partial}{\partial x_j}[(\mu + \frac{\mu_t}{\sigma \epsilon})\frac{\partial \epsilon}{\partial x_j}] + C_{1\epsilon}\frac{\epsilon}{k}(G_k + C_{3\epsilon}G_b) - C_{2\epsilon}\rho\frac{\epsilon^2}{k} + S_\epsilon \quad (2)$$

Due to the lack of space, we refer [4] and Ansys documentation[1] for further clarification of k-ε model.

In practice, models as k-ε are implemented through CFD software. Models evaluated in CFD result in very high accuracy, as shown, e.g., in [7]. The accuracy of the model is

[1]https://www.afs.enea.it/project/neptunius/docs/fluent/html/th/node58.htm

determined via two parameters: i) the sensitivity of the mesh η ii) the number of iterations ψ. The mesh should be sensitive enough (fine enough) to capture the geometry of the model in adequate level of details. However, too fine mesh also leads to a high computational complexity. Hence, there is a trade-off between the fine mesh leading to very accurate results at a cost of a high complexity while too coarse mesh with low complexity leads to inaccurate results. The time complexity of the FVM is $\mathcal{O}(\eta \log(\eta))$, where η is the mesh sensitivity characterizing the FVM environment. The FVM approximates values by a time evaluation in ψ discrete time steps, so the complexity becomes $\mathcal{O}(\eta \log(\eta)\psi)$ [8]. Hence, the model is very accurate, but in practice takes very long to converge.

PINNs were designed to surpass the complexity of CFD software with complexity of $\mathcal{O}(1)$. Unlike traditional neural networks, PINNs exploit the backpropagation algorithm and use the knowledge that each partial derivatives are used to propagate the error through the network. If we know that RANS are a set of PDEs, we can imagine each PDE as a part of backpropagation. The loss function of PINN is then composed of a loss function MSE_0, MSE_b and MSE_f [5]. MSE_0 is traditional mean squared error defined as $MSE_0 = \frac{1}{N_0}\sum_{i=1}^{N_0} |h(0, x_0^i) - h_0^i|^2$, where N_0 is the number of known datapoints h_0^i, gained from evaluation of the model in CFD software, $h(0, x_0^i)$ is the predicted value by neural network h for point x_0^i. MSE_b is the loss of periodic boundary and its definition can be found in [5]. $MSE_f = \frac{1}{N_f} = \sum_{i=1}^{N_f} |f(t_f^i, x_f^i)|^2$ is the loss of the results of the neural network gained at a set of collocation points x_f^i. Due to a page limit, we refer to [5] for a detailed description.

III. Experiments and results

We have used a simple geometry of 8 buildings with different heights encapsulated into a square-shaped wind tunnel. We have used academically available CFD Ansys Fluent to run the simulation using k-ε model. We have set the inlet velocity to 10ms^{-1} and the turbulent intensity to 2%. We select air with temperature 20^{o}C as the fluid. The convergence took 35 iteration and the time of run was 6 minutes. This is relatively fast simulation due to a simple geometry and low turbulent intensity. We set up the PINN network using NVDIA Modulus[2] framework based on the aneurysm example provided in the documentation. We use the same geometry mesh as in case of Ansys simulation. In Modulus, dynamic viscosity of the fluid has to be set manually, which is $\nu = 0.00001825$ for air with T =20^{o}C. Inlet velocity is set similarly to V=10ms^{-1}.

Modulus framework is designed to simplify work with PINNs. RANS are integrated into modulus and we provide an information about used differential equations as `NavierStokes(nu, rho, dim)`. In this code nu is the dynamic viscosity ν, rho is the air density $\rho = 1.204\text{kgm}^{-3}$ and dim represents the number of dimensions, in our case 3. In a line with official modulus documentation, we set up the architecture as a fully connected neural network with u, v and w velocity components as input and output of u, v, w and air pressure p. We set inlet and outlet similarly as in Ansys. We set no slip boundary conditions; the velocity is zero at boundary. We do not use MSE_0 as a loss function, as we want to get rid of using CFD software completely.

[2]https://developer.nvidia.com/modulus

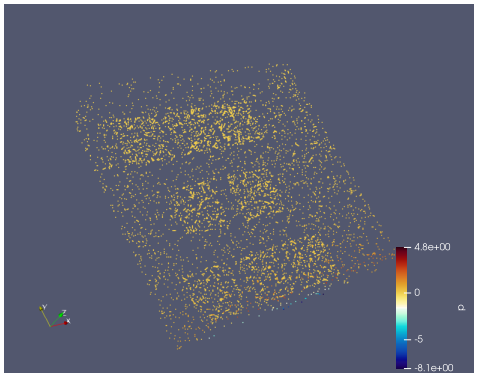


Fig. 1. A pressure on buildings caused by wind predicted by PINN visualized in paraview.

After evaluation of the model we compare the results of CFD and PINN. Using the MSE we conclude that the error for velocity prediction is 0.8ms-1, or approximately 10%. However, after training, the prediction of any value can be provided in milliseconds (16ms on average)), instead of minutes, as in case of Ansys. Also, we can predict the data in bulk as well, as visualized in 1.

IV. Conclusion and discussion

We have set up a physically informed neural network in NVIDIA Modulus framework and managed to predict wind with approximately 90% accuracy. This is relatively low accuracy for industrial use but makes a good starting point for deeper research in future. We have also proved that the prediction is super-fast in comparison to commercially used CFD software. A further future research has to be provided to find the best possible way to train PINN without governing data.

Acknowledgment

This work was supported by The Slovak Research and Development Agency project no. APVV-18-0214 and no. APVV SK-CZ-RD-21-0028.

References

[1] S. V. Patankar, *Numerical heat transfer and fluid flow*. CRC press, 2018.
[2] R. Eymard, T. Gallouët, and R. Herbin, "Finite volume methods," *Handbook of numerical analysis*, vol. 7, pp. 713–1018, 2000.
[3] T. Kajishima and K. Taira, "Reynolds-Averaged Navier–Stokes Equations," in *Computational Fluid Dynamics*. Springer, 2017, pp. 237–268.
[4] B. Mohammadi and O. Pironneau, *Analysis of the k-epsilon turbulence model*, 1993.
[5] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *Journal of Computational physics*, vol. 378, pp. 686–707, 2019.
[6] Y. Cao, R. Xu, and P. Jiang, "Physics-informed machine learning based rans turbulence modeling convection heat transfer of supercritical pressure fluid," *International Journal of Heat and Mass Transfer*, vol. 201, p. 123622, 2023.
[7] S. S. Das, J. C. Páscoa, M. Trancossi, and A. Dumas, "Computational fluid dynamic study on a novel propulsive system: Acheon and its integration with an unmanned aerial vehicle (uav)," *Journal of Aerospace Engineering*, vol. 29, no. 1, p. 04015015, 2016.
[8] J. Yang, T. Michael, S. Bhushan, A. Hanaoka, Z. Wang, and F. Stern, "Motion prediction using wall-resolved and wall-modeled approaches on a Cartesian grid," in *Proc. of the 28th Symposium on Naval Hydrodynamics (USA, Pasadena)*. Citeseer, 2010.

nnMixNet - when mixup augmentation meets nnU-Net architecture

[1]*Dávid Jozef Hreško,*
Supervisor: [2]*Peter Drotár*

[1,2]Department of Computers and Informatics, FEI TU of Košice, Slovak Republic

[1]david.jozef.hresko@tuke.sk, [2]peter.drotar@tuke.sk

***Abstract*—Application of deep learning models in domain of medical image segmentation has increased rapidly in recent years, mainly because of prominent results and capabilities to provide informative analysis in reasonable time and in an automated manner. However, reaching suitable level of generalization capabilities and achieving consistent precision over several independent domains is still challenging. To mitigate this issue, we propose novel combination of nnU-Net architecture and mixup augmentation, named nnMixNet. Proposed solution proved to be capable of effective generalization on several medical segmentation domains.**

***Keywords*—medical imaging, deep learning, mixup, image augmentation, semantic segmentation**

I. Introduction

For many years, researchers have been studying organ segmentation using medical imaging. Computed tomography, in particular, may offer precise information on organs and structures in the abdominal area. These data can be utilized for surgical planning, diagnosis, and a variety of clinical considerations. In recent years the contour delineation of anatomical and pathological features became crucial for applications such as visual assistance during surgery. While human demarcation is time consuming and tiresome, automated highly precise solutions are widely sought.

Convolutional neural networks are one of the most popular automated tools to perform medical image segmentation. Particularly, U-Net architecture [1] has become golden standard and still represents active topic for research and further modifications. Moreover, other preprocessing methods and various architectural improvements helped to popularize this set of neural networks. Considering all these recent advances, the organ segmentation appears to be almost solved problem [6]. However, adaptation and generalization of neural networks across multiple domains and similar datasets still represents challenge for automated segmentation and relies on manual analysis of specific characteristics.

To mitigate this issue, we propose unique combination of current state-of-the-art architecture for medical image segmentation, named nnU-Net [2] with novel mixup augmentation techniques. We assume that this combination can improve overall network performance and its generalization capability across different datasets and medical domains.

II. Methodology

This section describes concept of mixup augmentation and architecture of neural network used in our experiments.

A. Mixup augmentation

Besides the basic geometric augmentation techniques such as image rotation, flipping or cropping, which help to extend training data, there are other advanced methods such as mixup augmentations [3]. These methods rely on creation of synthetic data samples obtained from combination of two existing data samples. In practice, these combinations are created via linear interpolation, but combination process can vary between individual implementations and variants of mixup.

Concept of mixup augmentation was introduced in [4] and was mathematically described as:

$$\begin{aligned} \tilde{x} &= \lambda x_i + (1-\lambda) x_j \\ \tilde{y} &= \lambda y_i + (1-\lambda) y_j \end{aligned} \tag{1}$$

where x_i, x_j represents pair of input samples and y_i, y_j represents labels. Pairs (x_i, x_j), (y_i, y_j) are randomly sampled from training set and $\lambda \in [0, 1]$. Rate of interpolation between samples is controlled by hyperparameter α, which can be described as $\lambda \sim Beta(\alpha, \alpha), \alpha \in (0, \infty)$.

In recent years, multiple variants of mixup augmentation have been developed. In our research we focused on variant named manifold mixup [5]. This variant is very similar to original concept of mixup, however there are some key differences. The linear interpolation between pair of samples can be performed even on hidden layers, not only on the input one. With this approach there is additional information flow in network, thus overall capability to generalize is improved.

B. nnU-Net

nnU-Net [2] is a neural network architecture designed for medical image segmentation tasks. The name "nnU-Net" stands for "no-new-Net", meaning that the network is not fundamentally different from other existing networks, but rather a variation of the popular U-Net architecture that has been optimized for medical image segmentation.

The nnU-Net architecture is based on a 3D U-Net that is adapted to work with 2D, 2.5D or 3D medical image volumes, which makes it suitable for a variety of segmentation tasks in medical imaging, such as brain tumor segmentation, cardiac segmentation, and lung segmentation. The network is designed to handle datasets with a small number of annotated images, which is often the case in medical imaging, by using data augmentation techniques to generate synthetic images that can be used to train the network.

This architecture has achieved state-of-the-art results on several medical image segmentation challenges and has become a popular choice for researchers and practitioners in the field of medical imaging.

III. Medical datasets

In this section we provide short overview of medical datasets used to train and validate our model.

A. KiTS

KiTS (Kidney Tumor Segmentation) dataset [6] includes 300 publicly available CT scans with kidney, kidney tumor and cyst annotations. The patients in this dataset underwent partial or radical nephrectomy for suspected renal malignancy between 2010 and 2020 at either an M Health Fairview or Cleveland Clinic medical center. The kidney and tumor annotations were provided by medical students and laypeople (people who received no training other than a brief sheet with instructions) under the guidance of radiologists and urologic cancer surgeons.

B. KiPA

KiPA (Kidney Parsing) dataset [7] contains 130 3D abdominal CT images. The ground-truth corresponding to four classes, kidney, tumor, vein and artery, was determined by three medically trained experts and validated by experienced radiologist. All images were cropped to the same size of 150 × 150 × 200 to focus on the four structures of interests. The data itself were acquired by Siemens dual-source 64-slice CT scanner and the contrast media was injected during CT image acquisition. The further technical settings of CT are: X-ray tube current is 480 mA, B25F convolution kernel, exposure time equal to 500 ms and voltage 120KV. The slice thickness is 0.5mm/pixel and spacing of images is from 0.47 mm/pixel to 0.74 mm/pixel and 0.75 mm/pixel for z-direction, respectively.

IV. Experimental results

To validate our proposed solution we utilized two medical datasets. Dataset KiTS21 [8] was focused on segmentation of kidney and kidney structures from CT scans and KiPA22 dataset [9] was focused on segmentation of kidney, kidney structures, veins and arteries from CT scans. In case of KiTS21 dataset standard mixup was applied for nnU-Net. On the other hand, manifold mixup achieved better results during training stage on KiPA22 dataset, thus it was preferred choice over standard mixup.

TABLE I
Performance of nnMixNet on KITS21 validation data

network	Sørensen-Dice	Surface Dice	Tumor Dice
baseline nnU-Net	0.852	0.784	0.773
nnMixNet	0.877	0.793	0.795

Overall performance of our method was at first independently measured on KITS21 validation data. Sørensen-Dice and Surface Dice metric was determined for every HEC of every case of the validation set and averaged over each HEC. Additionally average Sørensen-Dice value on the Tumor HEC score was measured and was used as a tiebreaker in case of the same score. Achieved scores and comparison of baseline nnU-Net and our method can be seen in table I.

Second validation was performed on KiPA2022 data during test phase, where Dice metric was used to measure model performance. The obtained results and comparison with baseline nnU-Net can be seen in the table II.

TABLE II
Performance of proposed method on KiPA22 test data

	Kidney	Tumor	Vein	Artery
baseline nnU-Net	0.952	0.842	0.811	0.856
nnMixNet	0.957	0.88	0.835	0.878

V. Conclusion

In this paper we proposed multiple mixup enhancements of nnU-Net architecture. This solution was validated on two medical datasets. Obtained results proved that application of these methods can be beneficial in terms of overall network performance in domain of image segmentation. Additionally, we proved that combination of mixup augmentation and nnU-Net architecture does not bias to specific data characteristics and can be used as general solution to other segmentation tasks as well. In future work we plan to extend this solution with other augmentation blocks and to validate it on additional medical datasets.

Acknowledgment

The responsibility of the research content is on primary author and does not represent any view from funding authorities.

References

[1] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Springer*, 2015.
[2] F. Isensee, J. Petersen, A. Klein, D. Zimmerer, P. F. Jaeger, S. Kohl, J. Wasserthal, G. Koehler, T. Norajitra, S. Wirkert, and K. H. Maier-Hein, "nnu-net: Self-adapting framework for u-net-based medical image segmentation," *Nature methods*, vol. 18, 2020.
[3] K. Maharana, S. Mondal, and B. Nemade, "A review: Data pre-processing and data augmentation techniques," *Global Transitions Proceedings*, vol. 3, no. 1, pp. 91–99, 2022, international Conference on Intelligent Engineering Approach(ICIEA-2022). [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666285X22000565
[4] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," 2017. [Online]. Available: https://arxiv.org/abs/1710.09412
[5] V. Verma, A. Lamb, C. Beckham, A. Najafi, I. Mitliagkas, A. Courville, D. Lopez-Paz, and Y. Bengio, "Manifold mixup: Better representations by interpolating hidden states," 2018. [Online]. Available: https://arxiv.org/abs/1806.05236
[6] N. Heller, F. Isensee, K. H. Maier-Hein, X. Hou, C. Xie, F. Li, Y. Nan, G. Mu, Z. Lin, M. Han *et al.*, "The state of the art in kidney and kidney tumor segmentation in contrast-enhanced ct imaging: Results of the kits19 challenge," *arXiv preprint arXiv:1912.01054*, 2019.
[7] P. Shao, C. Qin, C. Yin, X. Meng, X. Ju, J. Li, Q. Lv, W. Zhang, and Z. Xu, "Laparoscopic partial nephrectomy with segmental renal artery clamping: Technique and clinical outcomes," *European Urology*, vol. 59, no. 5, pp. 849–855, 2011. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0302283810011395
[8] M. Gazda, P. Bugata, J. Gazda, D. Hubacek, D. J. Hresko, and P. Drotar, "Mixup augmentation for kidney and kidney tumor segmentation," 2021.
[9] D. J. Hresko, M. Kurej, J. Gazda, and P. Drotar, "Ensembled autoencoder regularization for multi-structure segmentation for kidney cancer treatment," 2022. [Online]. Available: https://arxiv.org/abs/2208.04007

Modeling of Supercapacitor for Hybrid Energy Storage System Applications

[1]*Dávid BODNÁR (2nd year)*
Supervisor: [2]František ĎUROVSKÝ

[1,2]Dept. of Electrical Engineering and Mechatronics, FEI TU of Košice, Slovak Republic

[1]david.bodnar@tuke.sk, [2]frantisek.durovsky@tuke.sk

***Abstract*—Supercapacitor technologies (SC) start to emerge in electric vehicle (EV) and stationary energy storage applications. They are often coupled with other energy storage technologies creating hybrid energy storage system (HESS). However, modeling diffusion processes in SC remains challenging. In this paper, a brief overview of possible supercapacitor models was made. Equivalent circuit model was adopted and created through Simscape Electrical in Matlab/Simulink. Parameter identification was made for 300 F SC cell, and simulation output was compared with measured data and showed high accuracy of the SC model. The created model can be used in SC management system as well as for designing purposes.**

***Keywords*—Supercapacitor, equivalent circuit model, Simscape Electrical, selfdischarging.**

I. Introduction

In recent years, supercapacitors gain more attention for their high power density, reliable operation, long lifetime, and low internal resistance [1]. There are three possible types of SC, electrical double-layer capacitor (EDLC), hybrid capacitors, and pseudo capacitors [2]. In this paper, EDLC will be referred to as SC.

In SC, both electrodes are produced from activated carbon. Their discharging characteristic is thus almost linear [3]. For modeling the electrical behavior of SC, electrochemical models, equivalent circuit models (ECM), or "Fractional order" models are used most frequently [4]. Artificial intelligence models such as neural networks can also be used [5]. SC modeling is a bit easier compared to battery modeling, since no chemical reactions take place in SC, and the SC voltage is approximately directly proportional to the SC charge. However, SC modeling still has its challenges, such as modeling diffusion processes, self-discharge, or thermal modeling.

II. Supercapacitor equivalent circuit model

ECM models are relatively widespread in SC modeling due to their simple implementation. The most used of them are shown in *Fig. 1*. The simplest ECM model SC is a resistor connected in series with a capacitor and one parallel resistor representing self-discharge (*Fig. 1* a)). In another way, the capacitor is connected in parallel to several parallel RC branches, which represent the diffusion processes in the SC cell (*Fig. 1* b)). RC pairs can also be connected in series with a capacitor (*Fig. 1* c)). The same ECM model can thus be used for both SC and batteries. The last method of ECM is shown in *Fig. 1* d), which was created to simulate the capacity of the porous electrode and the resistance of the electrolyte.

However, this method is quite difficult to identify the parameters. Self-discharge usually refers to two different mechanisms, namely ion diffusion and leakage current [4]. Ion diffusion is actually a redistribution of ions during charging and discharging because of the SC electrode porosity. The ion redistribution can be observed during a resting period as a voltage drop or increase. After a certain time of resting, the ions are evenly redistributed, and the SC voltage remains relatively stable. After charging, a voltage drop takes place, whereas after discharging, a voltage increase is observed in the SC cell [6]. This voltage increase is also referred to as self-charging [7]. These diffusion processes are responsible for the most significant changes in SC tension during resting time and can be mainly observed during the first few minutes. However, it takes hours for the SC to reach equilibrium. They are modeled using RC pairs as shown in Fig. 32 b) and c). These can have a very large effect on the overall voltage of the cell, especially for SCs with a smaller capacity.

Self-discharge also consists of a leakage current, since the electrolyte does not have infinite resistance. This current is very small, but during long-term storage of the SC, it can also significantly reduce the voltage or state of charge SC. The leakage current is usually modeled as a resistor connected in parallel with a relatively high value.

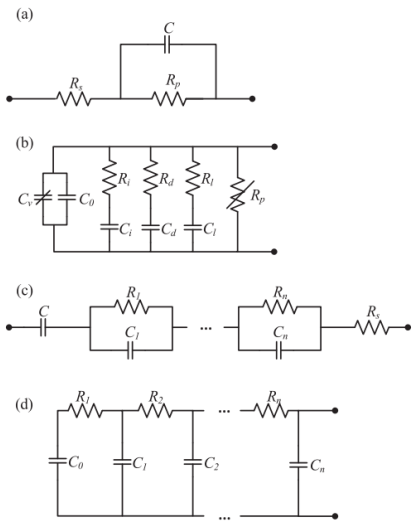


Fig. 1 ECM models of SC [4]

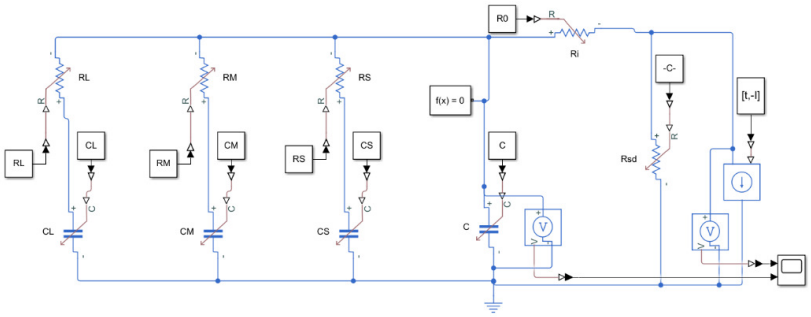


Fig. 2 ECM model SC in Simscape Electrical in MATLAB

III. SC MODEL IN MATLAB/SIMULINK

In order to create the SC model, SC cell type SCCY62B307SSB was selected for parameter identification. The SC cell has a nominal voltage of 2.7 V and a capacity of 300 F. All tests were performed on Arbin battery module tester, model LBT22043. We used Simscape Electrical in MATLAB for SC modeling (*Fig. 2*). We chose an ECM composed of three parallel-connected RC branches (*Fig. 1* b)) so that diffusion processes could also be modeled.

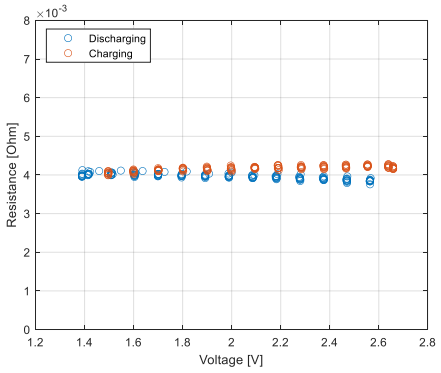


Fig. 3 Internal series resistance 300F SC depending on cell voltage

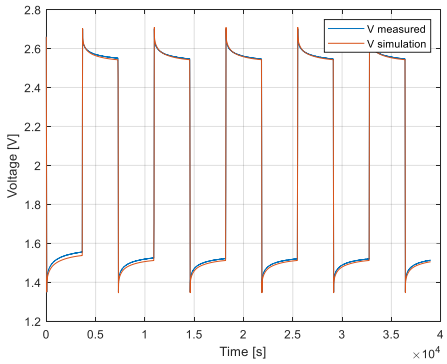


Fig. 4 Voltage response of 300F SC to discharge and charge pulses and comparison with SC model simulation results

The identification of parameters was made by Curve Fitting Toolbox in MATLAB, and a third-order ECM model was used. The result of the series internal resistance is shown in *Fig. 3*, depending on the SoC, or voltage changes only minimally and hovers around the value of 4 mΩ. Other parameters were verified according to the voltage response during the rest period, which lasted one hour (*Fig. 4*). The cell was alternately charged to 2.7 V and discharged to 1.35 V. In *Fig. 4*, the measured waveforms are compared with the simulation result, and only slight deviations were noted.

IV. CONCLUSION

In this paper, the SC model in MATLAB/Simulink was proposed. Simulation results show high accuracy of diffusion processes modeling. The created model can be used in the SC management system and for designing of HESS. SC self-discharging caused due to leakage current was not investigated in this paper, which requires further research. Moreover, the impact of temperature on self-discharging and diffusion was not investigated yet. This requires further experiments utilizing a temperature chamber. Extension of this work could be also modeling of the whole HESS including other energy storages like a battery pack, and necessary components of HESS, such as DC/DC converter. The efficiency of all employed components plays a crucial role in the overall efficiency of HESS. Therefore, it remains a subject of further study.

ACKNOWLEDGMENT

This work was supported by Slovak Research and Development Agency on the basis of Contract no. APVV-18-0436 and Contract no. APVV-15-0750.

REFERENCES

[1] A. Khaligh and Zhihao Li, “Battery, Ultracapacitor, Fuel Cell, and Hybrid Energy Storage Systems for Electric, Hybrid Electric, Fuel Cell, and Plug-In Hybrid Electric Vehicles: State of the Art,” *IEEE Trans Veh Technol*, vol. 59, no. 6, pp. 2806–2814, Jul. 2010, doi: 10.1109/TVT.2010.2047877.

[2] A. González, E. Goikolea, J. A. Barrena, and R. Mysyk, “Review on supercapacitors: Technologies and materials,” *Renewable and Sustainable Energy Reviews*, vol. 58, pp. 1189–1206, May 2016, doi: 10.1016/j.rser.2015.12.249.

[3] D. Slaifstein, F. M. Ibanez, and K. Siwek, “Supercapacitor Modeling: A System Identification Approach,” *IEEE Transactions on Energy Conversion*, pp. 1–11, 2022, doi: 10.1109/TEC.2022.3212617.

[4] L. Zhang, X. Hu, Z. Wang, F. Sun, and D. G. Dorrell, “A review of supercapacitor modeling, estimation, and applications: A control/management perspective,” *Renewable and Sustainable Energy Reviews*, vol. 81, pp. 1868–1878, Jan. 2018, doi: 10.1016/j.rser.2017.05.283.

[5] G. J. Adekoya, O. C. Adekoya, U. K. Ugo, E. R. Sadiku, Y. Hamam, and S. S. Ray, “A mini-review of artificial intelligence techniques for predicting the performance of supercapacitors,” *Mater Today Proc*, vol. 62, pp. S184–S188, 2022, doi: 10.1016/j.matpr.2022.05.079.

[6] J. Kowal *et al.*, “Detailed analysis of the self-discharge of supercapacitors,” *J Power Sources*, vol. 196, no. 1, pp. 573–579, Jan. 2011, doi: 10.1016/j.jpowsour.2009.12.028.

[7] J. J. Quintana, A. Ramos, M. Diaz, and I. Nuez, “Energy efficiency analysis as a function of the working voltages in supercapacitors,” *Energy*, vol. 230, Sep. 2021, doi: 10.1016/j.energy.2021.120689.

Overview of machine learning methods in the analysis of observation data

[1]*Lenka Kališková (1st year),*
Supervisor: [2]*Peter Butka*

[1,2]Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1]lenka.kaliskova@tuke.sk, [2]peter.butka@tuke.sk

***Abstract*—This paper provides an overview of multiple machine learning methods and additions that could enhance the proposed model and provide a great improvement in results. Explored methods include object detection, data augmentation, and physics principles that have been successfully applied in various fields. Our future goal is to transfer these methods to the research of the astrophysics domain with additional value.**

***Keywords*—Convolutional Neural Network, Generative Adversarial Network, Optimization of Hyperparameters, Image Segmentation, Next Frame Prediction, Object Detection, Physics Principles, YOLO**

I. Introduction

Methods for building and interpreting machine learning models in the analysis of observational data is a popular topic for research since artificial intelligence and machine learning are currently considered as peak topics in computer science. When discussing machine learning models, there are a great number of possibilities to choose from, but in this research, the main focus is centered on neural networks, specifically deep learning algorithms based on convolutional neural networks, with a primary focus on image-based data. We decided to work with the algorithms from the area of deep learning based on convolutional neural networks, like YOLO [1], with the use for different tasks like object detection [2] or image segmentation [3]. In addition to mentioned machine learning models, a powerful tool is also data augmentation [4], which can massively help with enlarging training inputs, and also optimization of hyperparameters [5] could greatly improve the overall performance of the model by finding the right settings of parameters. Lastly, worth mentioning is an approach that enhances machine learning models with basic and advanced physics principles and laws incorporated directly into the learning process. These models will work with images from the astronomical domain, where they could prove their power, and help with the classification and detection of new astronomical objects or the detection of an interesting phenomenon in the universe. For a better understanding of all mentioned terms, we recommend following the Deep Learning book [6].

Nowadays, there are many various physical phenomena not only on the Earth, but also in the space domain, which can provide us with this data in many forms, but most importantly for us, in the form of images. Now, we primarily focus on phenomena within the analysis of data from observations of the Sun and near-Earth domain, related to the so-called Space Weather Environment. The Sun is the star at the centre of our Solar System with a great number of physical phenomena, which have a significant impact not only on the Sun itself but also on the Earth. These phenomena can cause various events like coronal holes [7] or active regions, which we detect and explore. From the near-Earth domain, Transient Luminous Events (TLE) were discovered in the last decade of the twentieth century and are considered a greatly unexplored area [8]. These events are very rare and occur only during storms. TLE are described as optical events in the stratosphere, mesosphere, and at lower ionospheric altitudes. There are several various types of transient luminous events based on their shape, colour, or intensity. Another aspect of the near-Earth domain is the modelling of airglow, which is light that is observed especially during the night that originates in the high atmosphere of Earth, and that is associated with photochemical reactions of gases caused by solar radiation [9]. In all domain examples, our main effort is to support physicists with the appropriate algorithms for the detection, segmentation, and further analysis of selected events.

The paper is organized as follows. In the next section, we provide the state of the art regarding the use of machine learning for selected tasks and domains, including some new methods which we will see as option for future considerations, like the use of physics principles in learning. In conclusion, we provide some recommendations for future research.

II. State-of-the-art

As mentioned previously, our research covers different machine learning methods, which we use in astrophysics. Therefore, the current state is possible to divide into multiple sections for better clarity, as we provide an overview of works that used machine learning methods to address various issues and which we could use in our research to address astrophysics issues.

A. Object detection

Currently, one of the most popular rising machine learning models is You Only Look Once (YOLO), which fulfils an important role in the detection of objects within the image. Our research more specifically focuses on the detection of astronomical objects. Additionally, the latest version of YOLO provides image segmentation functionality.

First introduced in the paper [10] as a new approach to object detection. The authors were able to design an approach

consisting of only one neural network that detects objects in one single evaluation. Therefore, it is greatly used and helpful in various fields for the ability to detect an object in real-time, while it provides speed, high accuracy and level capabilities.

The study [11] focused on the identification of lung nodules in CT scans, which is considered a highly difficult task. To detect lung nodules, the authors used a convolutional neural network based on YOLO, as software based on the convolutional neural network already proven power in radiography. With using YOLO, the final results of classification are promising, as the achieved accuracy was 92.27%, which is a better result than without using YOLO. The Fig. 1 shows lung nodules detected by YOLO.

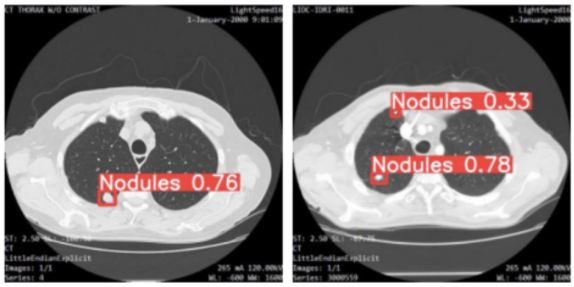


Fig. 1. Lung nodules detected by the YOLO model [12].

In [1], YOLO is used in a completely different field than the previous paper. Authors are focusing on the detection and classification of road damage. Keeping the roads maintain is important for a safe, economic and sustainable system of transport. By using machine learning, it is possible to ensure a safer and easier way to detect road damage. To detect road damage, the authors use a combination of Google Street View with YOLOv7 for six different countries. This combination proved to be efficient, as it achieves the highest F1 score at a level of 81.7% with the average F1 score calculated from the result of each country to be at the level of 74.1%. The Fig. 2 shows an example of damage detection from a road image.

Fig. 2. Road damage detected by the YOLO model [1].

To mention also usage of YOLO in the astronomical domain, there is currently research in progress, which focus on the detection of objects from the All-Sky Meteor Orbit System (system AMOS). The primary objective of system AMOS is to capture the entire sky to detect meteors but, during thunderstorms, the system is also able to detect transient luminous events (TLE). Since TLE are extremely sporadic, the best way to detect such events is through machine learning. To detect transient luminous events we use the mentioned YOLO detection network. Transient luminous events are considered as greatly unexplored and therefore are not many papers about this theme, which made it an exciting topic for further research. An exciting view on this topic is provided in [13]. The authors focused on the design and creation of a device for capturing transient luminous events. The creation of this device was important since there is currently no widespread process for the observation and processing of transient luminous events data. The created device consists of multiple parts, which are low gain, colour, and Complementary Metal Oxide Semiconductor (CMOS) camera with a bright, wide lens connected to a small power efficient single-board computer. In addition, the authors also provide software to manage a device. Software is able by itself to evaluate when is the great opportunity to capture transient luminous events and also with the use of neural networks is able to process and classify captured images. The Fig. 3 shows a photo captured by the created device.

Fig. 3. Phenomenom captured by the device created to detect transient luminous events [13].

B. Image segmentation

A similar approach to object detection, which is also a part of computer vision is image segmentation. The difference between these approaches is that image segmentation not only detects an object, but is also able to provide the shape of the detected object. The Fig. 4 shows the difference between object detection and image segmentation. Image segmentation can be accomplished through two primary approaches. The first approach involves traditional methods, such as clustering-based or edge-based techniques [3]. The second approach, which is of particular interest to us, utilizes machine learning methods.

Fig. 4. Difference between object detection and image segmentation [14].

General research about image segmentation is provided in the paper [15]. The authors examine multiple different machine learning models for image segmentation on the current most popular data sets to find the strengths of each model and to compare their performance, based on which future research could be established.

In [16], image segmentation is performed on images from the medical domain. Authors in this work examine three different architectures of convolutional neural networks for image segmentation, namely FCN, U-Net and DeepMedic.

C. Optimization of hyperparameters

Optimization of hyperparameters could provide a great improvement in the performance of the model. The two most common way to set hyperparameters of the model is hand-tuning or using optimization algorithms. Hand tuning is not considered an optimal method to optimize hyperparameters, because of all possible combinations that could exist, it could take weeks or months to find the optimal model based on a number of hyperparameters [5]. This is the reason, why optimization algorithms are primarily used for this task.

In [17], authors focused on optimising hyperparameters by using the algorithm Particle Swarm Optimization, which is a population-oriented metaheuristic algorithm. The authors decided to use this algorithm, as no paper before consider it as an option. The final results provide the efficiency of this choice, as it was able to discover optimal hyperparameters 81% faster than Grid Search, which evaluates every combination [18].

Another interesting approach was provided in [19]. To optimize a convolutional neural network, the authors created a variant of the genetic algorithm, which is based on the principle of choosing the best individuals from the population. Selected individuals then go through the process of mutation which leads to the creation of a new population [20]. The main difference is in the number of individuals, as the proposed solution has a variable number of individuals instead of a static number of individuals. This change leads to improvement, and therefore authors consider the use of variable individuals as promising.

D. Data augmentation

Expansion of data is possible to achieve with the use of a process called data augmentation. This process is primarily useful in scenarios when there is not enough data for training the chosen model [21]. The study [22] focuses on the detection and identification of galaxies with deep learning and data augmentation. As part of data augmentation, the authors created five different data sets with five different augmentation methods, e.g., changing colour scale, rotation or filtering of images. During the experiment authors also used a YOLO to detect and classify objects.

Another method for data augmentation is Generative Adversarial Network (GAN), which was introduced in [23]. In diploma thesis [24], author compared different variations of GAN for the augmentation of various galaxies' data. Augmentation with GAN provides images that prove to be more helpful than classic augmentation methods such as rotation or colour change. One of the used GAN variants was Data Augmentation GAN (DAGAN), which was first introduced in [25]. Authors augmented three different data sets and with the use of DAGAN, final results have improved from a range of 2.1% to 13% in comparison to classical augmentations methods. The Fig. 5 shows how images augmented by DAGAN look.

Fig. 5. Images created with the use of Data Augmentation GAN [25].

E. Physics principles for adaption in learning process

This section is dedicated to a combination of machine learning and physics principles, which indicates how these combinations could improve results and why it is an interesting approach to consider in our research.

Machine learning can occur in scenarios when training data are not particularly available. This is considered a great barrier to the implementation of models in different solutions since data availability is essential. In [26], the authors are dealing with the problem by proposing a solution based on a principle called the Fluctuation-Dissipation Theorem from statistical physics in combination with an artificial intelligence model for estimation response of our climate system to forcing and perturbation.

The study [27] is focusing on a combination of machine learning models with different physical laws, such as random points in the continuous space-time domain. Machine learning is considered an exciting substitute for solving problems which are difficult, but sometimes it is not considered an option because of absent data. This is where physics offers power since additional information could be acquired by already mentioned physical laws, but this is not the only advantage. With the use of physics, authors were able to improve neural networks which provide better accuracy, faster training and improved generalization.

Another combination of deep learning and physics is provided in [28]. Authors contemplate that machine learning is not able to challenge a more traditional physics and maths approach in the hardest and most complex challenges but

provide an example in form of a deep learning model, which gained knowledge from physics to predict the temperature of the sea surface. In the future, we would like to test the possibilities of such adaptions also for our selected domains and tasks.

F. Next Frame Prediction

Next-Frame Prediction is the captivating theme, which covers the possibility to predict future images by already known information from historical images. As this topic is still massively unexplored, it could be part of our research, since there is a lack of any approach to this problem in our domain.

The study [29] proposes two different architectures for next-frame prediction networks, which are sequence-to-one architecture and sequence-to-sequence architecture. These proposed architectures and design of the loss function are analysed and compared to find their positive and negative effects. Results, which were achieved show improvement to already existing methods for next-frame prediction. Improvement was not only achieved by proposed architectures but also by suitable design of loss functions. Based on this study, next-frame prediction is still only in the beginning and there is significant potential for the future. The study [30] proposes a method for next-frame prediction with the use of a special type of neural network called Convolutional Long-Short Term Memory that was introduced in [31]. The network is able from input in form of frames extract the features and forecast for further usage. During experiments, authors predict a sixth image in the row based on the previous five images. The Fig. 6 shows a predicted image. The final results were highly positive, as a proposed network was able to predict images effectively.

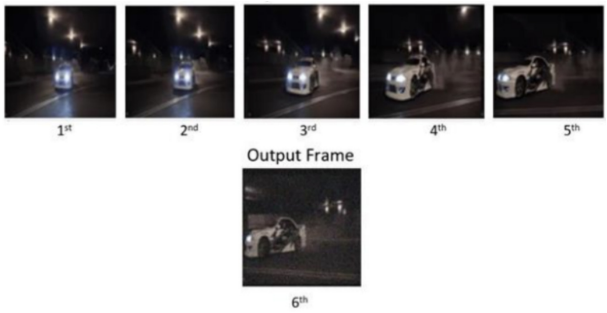


Fig. 6. Prediction of the sixth image in the row with next-frame prediction [30].

III. Conclusion

This paper focused on different types of machine learning methods and additions, which we are able to use in our research of the astrophysics domain. Our data are primarily in form of images and our baseline model is a convolutional neural network, which is already proved as the best option for this type of data. As mentioned, we explore other methods, which could in combination with our baseline model provide better results. Tasks we included are object detection and image segmentation, with additional techniques for improvements of neural network models like optimization of hyperparameters, data augmentation and physics principles. Our main goal is to apply mentioned methods in the astrophysics domain, where we will focus on using deep learning on images captured by system AMOS, namely meteors and transient luminous events, sun phenomena in the form of the solar corona objects (coronal holes and active regions) and solar flare ribbons. Special mention deserves a problem called next-frame prediction, as we will attempt to apply it for the modelling of airglow, something that was not done before.

Acknowledgment

This work was supported by the Slovak VEGA grant no. 1/0685/21.

References

[1] V. Pham, D. Nguyen, and C. Donan, "Road damages detection and classification with yolov7," *arXiv preprint arXiv:2211.00091*, 2022.
[2] Z. Zou, Z. Shi, Y. Guo, and J. Ye, "Object detection in 20 years: A survey," *arXiv preprint arXiv:1905.05055*, 2019.
[3] R. M. Haralick and L. G. Shapiro, "Image segmentation techniques," *Computer vision, graphics, and image processing*, vol. 29, no. 1, pp. 100–132, 1985.
[4] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of big data*, vol. 6, no. 1, pp. 1–48, 2019.
[5] L. Yang and A. Shami, "On hyperparameter optimization of machine learning algorithms: Theory and practice," *Neurocomputing*, vol. 415, pp. 295–316, 2020.
[6] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.
[7] S. R. Cranmer, "Coronal holes," *Living Reviews in Solar Physics*, vol. 6, pp. 1–66, 2009.
[8] E. Blanc, "Space observations of transient luminous events and associated emissions in the upper atmosphere above thunderstorm areas," *Comptes Rendus Geoscience*, vol. 342, no. 4-5, pp. 312–322, 2010.
[9] J. W. Chamberlain, *Physics of the Aurora and Airglow: International Geophysics Series, Vol. 2*. Elsevier, 2016, vol. 2.
[10] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.
[11] H. A. Ewaidat and Y. E. Brag, "Identification of lung nodules ct scan using yolov5 based on convolution neural network," *arXiv preprint arXiv:2301.02166*, 2022.
[12] K. Liu, "Stbi-yolo: A real-time object detection method for lung nodule recognition," *IEEE Access*, vol. 10, pp. 75 385–75 394, 2022.
[13] S. Amrich, Mackovjak, I. Strhárský, J. Baláž2, and M. Hančikovský.
[14] N. Murali, "Image classification vs semantic segmentation vs instance segmentation [online]," Published on 29. april 2023, https://nirmalamurali.medium.com/image-classification-vs-semantic-segmentation-vs-instance-segmentation-625c33a08d50 .
[15] S. Minaee, Y. Y. Boykov, F. Porikli, A. J. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," *IEEE transactions on pattern analysis and machine intelligence*, 2021.
[16] J. Bertels, D. Robben, R. Lemmens, and D. Vandermeulen, "Convolutional neural networks for medical image segmentation," *arXiv preprint arXiv:2211.09562*, 2022.
[17] B. Qolomany, M. Maabreh, A. Al-Fuqaha, A. Gupta, and D. Benhaddou, "Parameters optimization of deep learning models using particle swarm optimization," in *2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC)*, 2017, pp. 1285–1290.
[18] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *The Journal of Machine Learning Research*, vol. 13, pp. 281–305, 03 2012.
[19] X. Xiao, M. Yan, S. Basodi, C. Ji, and Y. Pan, "Efficient hyperparameter optimization in deep learning using a variable length genetic algorithm," 2020.
[20] T. Alam, S. Qamar, A. Dixit, and M. Benaida, "Genetic algorithm: Reviews, implementations, and applications," *arXiv preprint arXiv:2007.12673*, 2020.
[21] S. Yang, W. Xiao, M. Zhang, S. Guo, J. Zhao, and F. Shen, "Image data augmentation for deep learning: A survey," *arXiv preprint arXiv:2204.08610*, 2022.
[22] R. E. González, R. P. Muñoz, and C. A. Hernández, "Galaxy detection and identification using deep learning and data augmentation," 2018. [Online]. Available: https://arxiv.org/abs/1809.01691
[23] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2014.

[24] L. Kališková, "Využitie GAN sietí pre augmentáciu dát a zlepšenie klasifikácie dát," *Diploma thesis*, 2022.
[25] A. Antoniou, A. Storkey, and H. Edwards, "Data augmentation generative adversarial networks," 2017. [Online]. Available: https://arxiv.org/abs/1711.04340
[26] S. K. Kim, K. Ramea, S. R. Cachay, H. Hirasawa, S. Hazarika, D. Hingmire, P. Mitra, P. J. Rasch, and H. A. Singh, "Climate intervention analysis using ai model guided by statistical physics principles," to be published. [Online]. Available: https://arxiv.org/abs/2302.03258
[27] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, "Physics-informed machine learning," *Nature Reviews Physics*, vol. 3, no. 6, pp. 422–440, 2021.
[28] E. de Bezenac, A. Pajot, and P. Gallinari, "Deep learning for physical processes: Incorporating prior scientific knowledge," 2017. [Online]. Available: https://arxiv.org/abs/1711.07970
[29] Y. Zhou, H. Dong, and A. El Saddik, "Deep learning in next-frame prediction: A benchmark review," *IEEE Access*, vol. 8, pp. 69 273–69 283, 2020.
[30] P. Desai, C. Sujatha, S. Chakraborty, S. Ansuman, S. Bhandari, and S. Kardiguddi, *Next frame prediction using ConvLSTM*, 2022, vol. 2161, no. 1.
[31] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-c. Woo, "Convolutional lstm network: A machine learning approach for precipitation nowcasting," *Advances in neural information processing systems*, vol. 28, 2015.

Flexible Ag_2S Based Thermoelectrics for Use in Thermoelectric Generators

[1]Gabriela HRICKOVÁ (2nd year)
Supervisor: [2]Juraj ĎURIŠIN

[1,2]Dept. of Technologies in Electronics, FEI TU of Kosice, Slovak Republic

[1]gabriela.hrickova@tuke.sk, [2]juraj.durisin@tuke.sk

***Abstract* — The paper deals with new thermoelectric materials based on silver sulfide (Ag_2S). It explains production and discusses electric, and thermoelectric properties of Ag_2S-based flexible thermoelectric materials potentially applicable in wearable electronics.**

***Keywords* — thermoelectric material, Ag_2S, electrical conductivity, Seebeck coefficient, PF, X-ray diffraction.**

I. Introduction

The use of thermoelectric materials is not limited to temperature measurement. They can also be used for energy conversion, such as in thermoelectric generators, which convert heat energy into electrical energy. This is done by utilization the Seebeck effect, which is the conversion of a temperature difference into an electric voltage. Thermoelectric materials are also used in cooling applications, such as in thermoelectric coolers, which use the Peltier effect to cool objects. In this case, a temperature difference is created by applying an electric current to the material.

Silver sulfide (Ag_2S) is a promising thermoelectric material. Ag_2S without dopants does not show suitable properties to be used in a thermoelectric generator. In the case of Ag_2S, dopants can be used to increase its Seebeck coefficient and power factor and to improve its performance as an energy transporting material [1], [2].

II. Production and Measurements

To obtain powders of pure elements, i.e. Ag, S and the dopands Se and Te is the first step in the production of Ag2S based thermoelectric materials. The powders should have a purity of 99.99% and be free of any contaminants. The next step is to mix the powders in the desired proportions to obtain the required thermoelectric materials composition. The dopants, Se and Te, are both non-metallic elements. Both elements are in general used as dopants in semiconductor materials to improve their electrical properties. Se and Te are both effective at increasing the conductivity of semiconductor materials, and they can also be used to adjust the band gap of the material. The dopants help to improve the electrical and thermal properties of the sample.

Finally, the samples have to be tested to determine their thermoelectric properties. This can be done by measuring the Seebeck coefficient, electrical conductivity, and thermal conductivity of the sample [3]–[5].

A. Samples preparation

The first step in producing a thermoelectric material based on Ag_2S is to synthesize the material. This can be done by a variety of methods, we use synthesis in the solid state (annealing) from powders of input elements (Ag, S, Se, Te). We produced 5 samples with different compositions. The samples composition is the following: Ag_2S, $Ag_2Se_{0.4}Te_{0.1}S_{0.5}$; $Ag_2Se_{0.3}Te_{0.2}S_{0.5}$; $Ag_2Se_{0.2}Te_{0.3}S_{0.5}$ a $Ag_2Se_{0.1}Te_{0.4}S_{0.5}$. The powder mixture for each sample was always well mixed before annealing to produce a homogeneous composition. The powder mixture was then synthesized in the annealing container in an argon atmosphere at 160°C for 10 hours in an annealing oven. The annealed powder samples were then turned into solid rods by the rapid cooling (10^6 K^{-1}) of the powders melt. To obtain such a rapid cooling, the melts were injected in a copper form in an argon atmosphere. Thus, for the evaluation of the Ag_2S based thermoelectrics, only the rapidly cooled samples were used because the rod shaped samples are compact contrary to the incompact powders. The only exception was the evaluation by X-ray diffraction, where both the powder and rod-shaped samples were analyzed.

B. Seebeck coefficient, conductivity and power factor of Ag_2S

In this measurement, the Seebeck coefficient α was measured near room temperature, with a temperature difference between the hot and cold ends of the rod sample being $\Delta T = 17$ K. In Fig. 1 it can see that with increasing content of Te, the Seebeck coefficient α increases and then decreases (for absolute values of α the trend is vice versa).

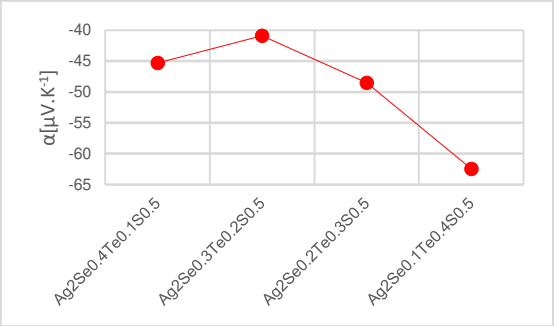


Fig. 1 Seebeck coefficient α vs. various compositions of Ag_2S based rod samples

Electric conductivity σ was measured by 4 wire Kelvin method. In Fig. 2, we can see the same situation as with the

Seebeck coefficient α. With the increasing content of Te the conductivity σ of the studied materials increases and then decreases. The conductivity σ of the rod samples with dopants is higher than for the pure rod sample without dopants. The conductivity σ of pure Ag_2S is 0.148 $S.m^{-1}$.

As we can see in Fig. 1 and Fig. 2, that increasing the Seebeck coefficient α causes decreasing the electric conductivity σ of materials.

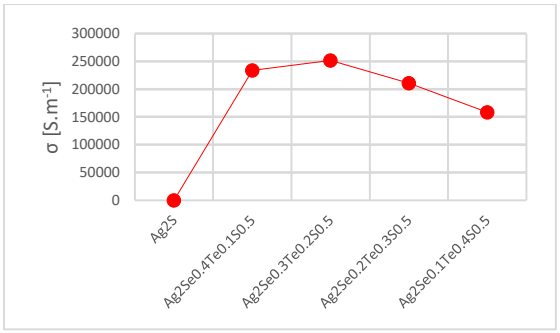


Fig. 2 Electrical conductivity vs. various compositions of Ag_2S based rod samples

In Fig. 3, the power factor *PF* is depicted for is depicted each rod sample. The power factor *PF* increases with the increasing dopant concentration (Te). The highest power factor *PF* is obtained for Te = 0.4. The power factor *PF* of Ag_2S with dopants is higher than that of pure Ag_2S, which indicates that the dopants can improve the power factor *PF* of the material.

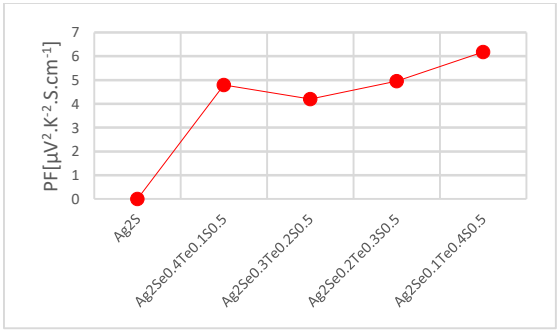


Fig. 3 Power factor PF vs. various compositions of Ag_2S based rod samples

C. X-ray diffraction

The samples (rods, powders) were exposed to X-ray radiation with a wavelength of λ = 1.54187 Å during the measurements. The diffracted beams were detected using a position-sensitive detector in the Bragg angle range 2θ = 10-140°. For a better clarity, diffraction patterns only in the range of 2θ = 25-55° are shown in the following figures.

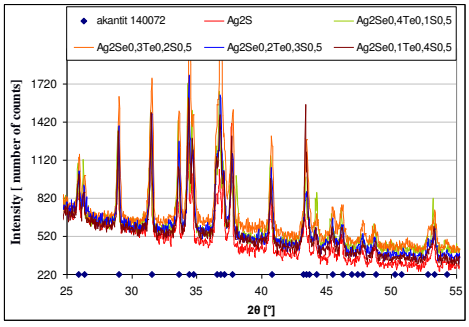


Fig. 4 X-ray diffraction patterns of powder Ag_2S based samples

The X-ray diffraction measurements (Fig. 4) showed that the pure and dopant Ag_2S annealed powder samples are crystalline, their structure corresponds to α-Ag_2S crystal lattice. The diffraction peaks were observed at angles corresponding to the Bragg's law. The peaks were sharp and well-defined, indicating that the peak positions for each sample are identical to the peak positions given in the database for α-Ag_2S. The addition of dopants (Se, Te) to α-Ag_2S does not cause a change of the crystalline system of the α-Ag_2S based samples. Fig. 5 shows diffraction patterns of the α-Ag_2S-based powder samples obtained from crushed rods produced by the rapid-cooling casting method, hence without annealing. Compared to previous diffraction patterns of the annealed powder samples (Fig. 4), the diffraction patterns show here lower relative intensities of the Bragg peaks.

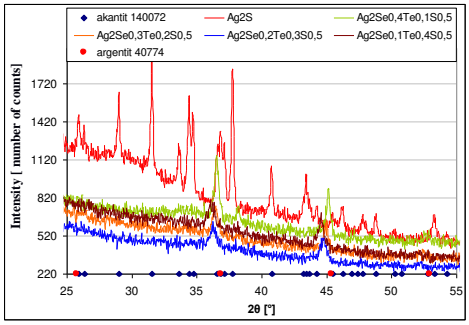


Fig. 5 X-ray diffraction patterns of Ag_2S-based rod samples

The decrease of the intensity is particularly pronounced for α-Ag_2S with the dopants (Se, Te). The decrease of the intensity indicates that the rapidly cooled samples show a lower proportion of crystalline phase than the annealed samples, thus exhibiting an amorphous structure in addition to the α-Ag_2S crystalline structure, this is particularly evident for α-Ag_2S with dopants (Se, Te). The partially amorphous nature of the rapidly cooled α-Ag_2S based rod samples gives an assumption of their better thermoelectric properties compared to their fully crystalline counterparts.

III. CONCLUSION

The addition of Se and Te dopants increases the Seebeck coefficient α and the power factor *PF* of the α-Ag_2S-based thermoelectrics. The X-ray measurements showed that the rapid cooling casting method modifies the dopant α-Ag_2S crystalline structure towards more amorphous structure. The preparation of α-Ag_2S-based samples by the rapid cooling casting method could be a suitable way to achieve an increase in the value of the figure of merit *Z.T* and to improve the thermoelectric properties of Ag_2S-based materials.

REFERENCES

[1] J. Liang *et al.*, 'Flexible thermoelectrics: from silver chalcogenides to full-inorganic devices', *Energy Environ. Sci.*, vol. 12, no. 10, pp. 2983–2990, 2019, doi: 10.1039/C9EE01777A.

[2] L. Li *et al.*, 'Study the effect of alloying on the phase transition behavior and thermoelectric properties of Ag2S', *Journal of Alloys and Compounds*, vol. 886, p. 161241, Dec. 2021, doi: 10.1016/j.jallcom.2021.161241.

[3] H. Wang *et al.*, 'High-Pressure Rapid Preparation of High-Performance Binary Silver Sulfide Thermoelectric Materials', *ACS Appl. Energy Mater.*, vol. 4, no. 2, pp. 1610–1618, Feb. 2021, doi: 10.1021/acsaem.0c02810.

[4] H. J. Goldsmid, *The Physics of Thermoelectric Energy Conversion*. IOP Publishing, 2017. doi: 10.1088/978-1-6817-4641-8.

[5] G. A. Pal'yanova, K. V. Chudnenko, and T. V. Zhuravkova, 'Thermodynamic properties of solid solutions in the system Ag2S–Ag2Se', *Thermochimica Acta*, vol. 575, pp. 90–96, Jan. 2014, doi: 10.1016/j.tca.2013.10.018.

Overview of datasets used for grammar error correction in NLP

[1]*Maroš HARAHUS (2nd year),*
Supervisor: [2]*Matúš PLEVA*

[1,2]Department of Electronics and Multimedia Communications, Faculty of Electrical Engineering and Informatics
Technical University of Kosice, Letná 9, 042 00 Košice, Slovak Republic, Tel. +421 55 602 2298

[1]maros.harahus@tuke.sk, [2]matus.pleva@tuke.sk

***Abstract*—The field of natural language processing has made significant progress in recent years, with an increasing focus on the development of algorithms for detecting and correcting grammatical errors in the text. A key component of this effort is the creation of high-quality annotated corpora that serve as a source of training data for these algorithms. These corpora are typically created by manually annotating text samples with information about grammatical errors and their correction. The quality of the corpus directly affects the accuracy and performance of the error correction models that are built on top of it. This paper provides an overview of the most common corpora in the development of grammar error correction systems. The paper concludes by highlighting the current challenges in this area and possible future research directions.**

***Keywords*—grammar error correction, dataset, corpus, NLP**

I. Introduction

Grammar error correction is crucial for applications like language learning, machine translation, and text-to-speech systems. To develop effective error correction algorithms, we need large annotated corpora for model training and evaluation. A high-quality grammar error correction corpus should include numerous text samples, diverse genres and language styles, and a comprehensive set of grammatical errors.

II. Datasets

Quality models in NLP depend on the quality of the data. Thousands to hundreds of thousands of data help to improve the training of language models in NLP. Preparing good-quality data sometimes takes hundreds of hours. For correcting grammatical errors, there are comparatively less data available than, for example, in machine translation. In this section, we introduce the most commonly used datasets. We will first describe English datasets and later move on to datasets in other languages.

A. English Datasets

Cambridge Learner Corpus First Certificate in English: (CLC-FCE) corpus was first described in ANewDataset and Method for Automatically Grading ESOL Texts [1]. This corpus consists of the Cambridge Learner Corpus (CLC). This corpus consists of 1,244 scripts (531 thousand words). All scripts are anonymized and annotated using XML. Each script also contains meta-data, which stores information about the question assignments, the candidate's grades, his/her native language, and age [1]. The FCE corpus has been manually annotated and contains 80 error types. It contains 30,995 pairs of parallel sentences that we can use to train our model [2]. The data is divided into standard training, development and test sets [3]. FCE is also commonly used to train models for automatic grammatical error correction [4] [5].

Lang-8: The Lang-8 corpus is referred to as the Learning English corpus. According to the paper [6], the authors created a large corpus with temporal/aspect errors from the Lang-8 social network. This "Social Network " is designed for students learning English as a secondary language. Students post their texts which are corrected by native English speakers. Lang-8 contains 509,116 pairs of sentences written by Japanese students [7]. The corpus is commonly used as a resource for training and evaluating grammar error correction models, which is the task of automatically detecting and correcting grammatical errors in written text.

JFLEG: The Johns Hopkins Fluency-Extended GUG (JFLEG) corpus is created by artificially inserting errors into fluent sentences and then giving the models a task to correct these errors. According to the authors of the corpus described in this paper [8], the corpus contains 1,501 sentences (28.1 thousand words) divided approximately equally into development and test sets. The sentences were randomly selected from essays written by students of English for whom English is a secondary language. Each sentence was annotated a total of 4 times, resulting in 4 sets of parallel reference annotations. This corpus is an extended version of the previously published JHU GUG (FLuency-Extended GUG-Grammar-Ungrammar).

W&I+LOCNESS: The Writing and Informatics + Linguistically Annotated Corpus of Non-native English Speaking Students (Write & Improve (W&I) and LOCNESS) is a grammar error correction corpus used to evaluate the performance of natural language processing (NLP) models in identifying and correcting grammatical errors in the text. The datasets consist of a total of 3,600 essays containing different types of grammatical errors [9]. The corpus is divided into training and test data. The training and test data were annotated by different annotators. The training data was annotated by only one annotator while the test data was annotated by 5 annotators. Edits were explicitly defined but not manually classified, so error types were added automatically using the ERRANT tool - an open-source grammar error correction toolkit [10].

B. Datasets in other languages

Czech: *AKCES-GEC* - Czech Language Acquisition Corpora is a grammar error correction corpus for the Czech language. According to the authors [11], the corpus is divided into training development and test data. Each dataset is divided into foreign and Roma learners. The development and test datasets are further divided into Slavic and non-Slavic speaking students. The essays and exam-style scripts come from the Corpus for Learners of Czech as a Second Language. AKCES-GEC corpus using annotation of three levels. Level 0 was used as source texts, level 2, as corrected texts, and error corrections were created according to manual alignment while retaining the error annotations [11]. *GECCC* - The Grammar Error Corpus for Czech (GECCC) is a corpus of the written text in the Czech language that contains annotated grammatical, spelling, and punctuation errors. It was created in 2022 [12]. The corpus consists of 83 thousand sentences (949 thousand words). The corpus was manually annotated by 5 annotators. The GECCC corpus is an extension of the AKCES-GEC corpus. Compared to the previous AKCES-GEC corpus, the GECCC corpus also contains formal texts written by native Czech primary and secondary school students, as well as informal web discussions on Facebook and Czech news websites.

German: *Falko* - The Falko corpus [13] is a German corpus for correcting grammatical errors. The corpus consists of 24 thousand sentences (381 thousand words) written by learners. It consists of textual data that has been annotated with grammatical errors. *MERLIN* - The MERLIN [13] corpus consists of standardized scripts for the German language examinations. The tests were not manually annotated but adapted using the ERRANT tool.

Ukrainian: Ukrainian Automated Grammar Error Correction Corpus (UA-GEC)[14] It is a set of textual data that has been annotated with grammatical errors and is intended for the development and evaluation of systems for grammatical error correction. It consists of 20 thousand sentences (329 thousand words) written by almost 500 authors. The edits were explicitly annotated and classified according to a scheme of 4 error types: grammar, spelling, punctuation, or fluency.

Dataset	Size (in words)	No. of Sentences
CLC-FCE	531,000	30,995
Lang-8	-	509,116
JFLEG	28,100	1,501
W&I+LOCNESS	-	3,600
AKCES-GEC	-	-
GECCC	949,000	83,000
Falko	381,000	24,000
MERLIN	-	-
UA-GEC	329,000	20,000

TABLE I
DATASET SIZE AND NUMBER OF SENTENCES

III. CONCLUSION

The described datasets are the most used datasets in model training for grammar error correction. There are many other datasets e.g. NUCLE, Gigaword etc. that have not been mentioned in this article. In conclusion, the production of high-quality annotated corpora plays a key role in the development of effective algorithms for grammatical error correction. A well-constructed corpus provides the training data needed to train and evaluate these algorithms, and its quality directly affects the accuracy and performance of error correction models. We want to create a high-quality corpus in the Slovak language for grammar error correction. We want to download grammatically correct sentences from various websites that manage the Slovak corpus. We assume that these websites should not contain text with grammatical errors. We want to use the *Spacy* tool to annotate the text. We plan to collect grammatically incorrect sentences from elementary school students who write essays and dictations or use a neural network to generate grammatical errors in sentences.

ACKNOWLEDGMENT

This work was supported by the Slovak Research and Development Agency under Grant APVV-SK-TW-21-0002;, and by the Faculty of Electrical Engineering and Informatics, TU Košice under the Grant FEI-2023-95, and under Grants VEGA 2/0165/21 and KEGA 055TUKE-4/2023 funded by the Ministry of Education, Science, Research and Sport of the Slovak Republic.

REFERENCES

[1] H. Yannakoudakis, T. Briscoe, and B. Medlock, "A new dataset and method for automatically grading esol texts," in *Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies*, 2011, pp. 180–189.

[2] Z. Yuan and T. Briscoe, "Grammatical error correction using neural machine translation," in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2016, pp. 380–386.

[3] Council of Europe. Council for Cultural Cooperation. Education Committee. Modern Languages Division, *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press, 2001.

[4] M. Rei and H. Yannakoudakis, "Compositional sequence labeling models for error detection in learner writing," *arXiv preprint arXiv:1607.06153*, 2016.

[5] S. Bell, H. Yannakoudakis, and M. Rei, "Context is key: Grammatical error detection with contextual word representations," *arXiv preprint arXiv:1906.06593*, 2019.

[6] T. Tajiri, M. Komachi, and Y. Matsumoto, "Tense and aspect error correction for esl learners using global context," in *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 2012, pp. 198–202.

[7] T. Mizumoto, Y. Hayashibe, M. Komachi, M. Nagata, and Y. Matsumoto, "The effect of learner corpus size in grammatical error correction of esl writings," in *Proceedings of COLING 2012: Posters*, 2012, pp. 863–872.

[8] C. Napoles, K. Sakaguchi, and J. Tetreault, "Jfleg: A fluency corpus and benchmark for grammatical error correction," *arXiv preprint arXiv:1702.04066*, 2017.

[9] C. Bryant, M. Felice, Ø. E. Andersen, and T. Briscoe, "The bea-2019 shared task on grammatical error correction," in *Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications*, 2019, pp. 52–75.

[10] C. Bryant, M. Felice, and E. Briscoe, "Automatic annotation and evaluation of error types for grammatical error correction," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*. ACL, 2017, pp. 793–805.

[11] J. Náplava and M. Straka, "Grammatical error correction in low-resource scenarios," *arXiv preprint arXiv:1910.00353*, 2019.

[12] J. Náplava, M. Straka, J. Straková, and A. Rosen, "Czech Grammar Error Correction with a Large and Diverse Corpus," *Transactions of the Association for Computational Linguistics*, vol. 10, pp. 452–467, 04 2022. [Online]. Available: https://doi.org/10.1162/tacl_a_00470

[13] A. Boyd, "Using wikipedia edits in low resource grammatical error correction," in *Proceedings of the 2018 EMNLP Workshop W-NUT: The 4th Workshop on Noisy User-generated Text*, 2018, pp. 79–84.

[14] O. Syvokon and O. Nahorna, "Ua-gec: Grammatical error correction and fluency corpus for the ukrainian language," *arXiv preprint arXiv:2103.16997*, 2021.

Accelerating Computer Vision Tasks via Vehicular Edge Computing

[1]*Róbert RAUCH (2nd year),*
Supervisor: [2]*Juraj GAZDA*

[1,2]Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

[1]robert.rauch@tuke.sk, [2]juraj.gazda@tuke.sk

***Abstract*—Deep learning is a critical technology for applications such as autonomous driving that requires speed, reliability, and accuracy. To accelerate the inference of increasingly complex deep neural networks (DNNs), novel approaches such as early exiting (EE) and split computing (SC) are emerging. EE allows the inference of neural networks to exit early if confidence is high enough, while SC splits the neural network into head and tail sections that are sequentially executed on device and server. However, to enable EE and SC, more research needs to be done to find optimal exit and split points. To address these issues, we first added an autoencoder to our DNN and then developed a framework that chooses split and early exit confidence thresholds based on linear regression models. Our approach is faster and achieves higher accuracy when compared to the state-of-the-art Edgent framework.**

***Keywords*—Computer Vision, Early Exiting, Split Computing, Vehicular Edge Computing**

I. Introduction

The idea of moving computational and storage capabilities outside of the device has been around for a long time. Computation and storage on a remote cloud server, also known as Cloud Computing (CC), has been motivated on a large scale by the rising number of Internet of Things (IoT) devices, with some researchers [1] predicting as many as 500 billion devices by 2030. Since clouds are centralized, latency to the server is often affected by increased traffic, geological location and other factors. Extensive research [1] has been conducted into how improving mobile networking technologies can decrease latency to these cloud servers. Such research is often focused even beyond current 5G technologies, such as 6G.

With the rise of 5G mobile networks, the European Telecommunications Standards Institute (ETSI) introduced a new paradigm called Mobile Edge Computing (MEC) [2] to tackle the issue with latency and increased traffic on cloud servers. The principle of MEC lies in moving computational and storage servers closer to the user. These servers are often placed at the edge of the network, therefore often at the base stations of a mobile network.

Usually users, also called edge devices or User Equipment (UE), are some IoT devices, phones or other devices with restrictive computational power and energy limitations. In our research, we presume UE to be autonomous vehicles. Because of this, we will also be thinking of our paradigm as Vehicular Edge Computing (VEC). Compared to MEC, this paradigm is required to take into account high speed of UE. VEC is also often used with specific applications [3] in mind, such as Global Positioning System (GPS), health monitoring systems, crash detection systems and others.

In VEC, UE often communicates between each other to share information. Depending on hierarchy of communication we divide VEC into Vehicle to Vehicle (V2V), Vehicle to Infrastructure (V2I) and Vehicle to Everything (V2X).

Main contribution of this paper lies in implementation of our own framework, which compared to the state-of-the-art Edgent [4] achieves lower latency and higher accuracy. This improvement is thanks to the linear regression model that determines confidence thresholds of each early exit.

II. Related Works

In this paper, we will be using early exiting and split computing to improve the execution latency of computer vision tasks. In 2016, Teerapittayanon et al. [5] introduced a way to lower execution latency of classification task by introducing early exits into the neural network. Inference of the neural network went through each of the exits and got confidence of that exit to determine whether its confidence is sufficient to predict correct target value. Using this method not only did they improve execution time but also improved accuracy in some cases. To train their branched models they firstly got prediction from each exit and then they used backpropagation with joint loss function from each exit. Another way we can train early exits, as proposed by Laskaridis et al. in [6], is to train backbone of the model first and then train each of the exits separately with backbone being frozen.

Another obvious way we can improve latency of task execution is to offload this task to the much more computationally capable server. We can, however, also split this execution between UE and server. Such an approach was proposed by Kang et al. in [7]. They did, however, find that best placement for the split is either first or last layer. In other words either sending entire task to the server or executing entire task on UE. To make split points more convenient for offloading, Eshratifar et. al. introduced BottleNet framework in [8], which compresses the intermediate data.

These two methods are often used simultaneously to create early exit execution and split strategy to improve task execution. One such example is Edgent, proposed by Li et al. in [4], where they use a simple algorithm to determine execution and split strategy.

III. Computer Vision Task

We will be using the same model for our classification as our last research in [9]. Compared to this research, we added a split after each of the VGG-16 blocks. Each of these splits also has an autoencoder, which encodes the output of the intermediate layer into a smaller size and then these data can be transferred to the server, where they are decoded. All possible compression with our autoencoder are shown in Table I with their space saving, accuracy impact and byte size of the encoder output. Values shown in bold are the ones that were used in our research. An image of our model architecture is provided in Appendix A.

For our task execution we will be using a simple equation to calculate latency of task exeuction L:

$$L = \frac{\mathcal{I}^{device}}{\eta^{device}} + \frac{\mathcal{I}^{server}}{\eta^{server}} + \frac{D}{B}, \tag{1}$$

where $\mathcal{I}$ is number of instructions, that need to be executed either on device or on server, η is computational power of a CPU either on device or on server, D is byte size, that needs to be sent to the server and B is current bandwidth to the server.

IV. Framework for Choosing Split and Execution Strategy

Our framework was heavily inspired by the Edgent framework from [4]. More specifically, their optimizer for static environment algorithm. We, however, don't loop through all exits but rather only loop through all possible splits. We then predict execution time on device and on server using our $split\_regressor$. This regressor is similar to the Edgent regressor with the exception that we don't predict execution based on the input size but rather based on the instructions I^{device} that need to execute to split (these are executed on device) and instructions I^{server} that need to execute from the split to the end of the model. This change was motivated by the fact that we don't presume that we have a split after each layer and we can easily determine the number of instructions that need to be executed based on the model structure. In our case, we measured instructions using *perf-stat* command on Linux. Instructions also include instructions for encoder and decoder. We save each latency for a split in a $latencies$ list by summing predicted execution time on device ED, on server ES and time needed to offload the task, represented as size in bytes D divided by bandwidth B. We can determine best split point by finding minimum value in our list of latencies. The biggest difference between our approach and Edgent is that we don't loop this through all the exits. This was motivated because time difference between exits was not sufficient for algorithm to effectively choose different exit, only in a small range of latency requirement values. Therefore their model for most cases evaluated that best strategy is to go through entire network and offload right at the beginning or don't offload at all if bandwidth is too low. This can have impact not only on speed of execution but also on accuracy, since as proven in [5] sometimes it is beneficial to exit at early exits, because later exits can be overconfident on wrong classes. In our case, rather than determining which exit to choose, we determine confidence thresholds for each exits (except the last one as we have to exit at this point). We can then execute our model, while going to each exit and if exit has confidence higher than this threshold we exit at this exit. We determine these confidence thresholds by another regression model, which takes latency requirement, which instead of latency in milliseconds we input ratio of images we would prefer to exit at each exit and split we found.

Algorithm 1 Strategy decision

for $i \leftarrow 0..n$ **do**
  $ED_i \leftarrow split\_regressor(\mathcal{I}_i^{device})$
  $ES_i \leftarrow split\_regressor(\mathcal{I}_i^{server})$
  $latencies \leftarrow ED_i + ES_i + D_i/B$
end for
$split \leftarrow argmin\{latencies\}$
$thresholds \leftarrow exit\_regressor(Latency, split)$
Return $split, thresholds$

V. Evaluation

To evaluate our model we used Belgium 4G/LTE dataset[1] with measured bandwidths to the server with different vehicles such as buses, tram etc. Figure 1 shows, that our model outperforms Edgent by a small amount with speed of execution represented in a form of frames per second (fps) and by accuracy. Reason for only a small amount of improvement in fps is due to the processing capabilities of both server and device, where we chosen their processing power as 2.000.000 MIPS and 10.000 MIPS, respectively. With most instructions that needs to be processed being about 540 million. This means, that biggest bottleneck was in offloading task to the server.

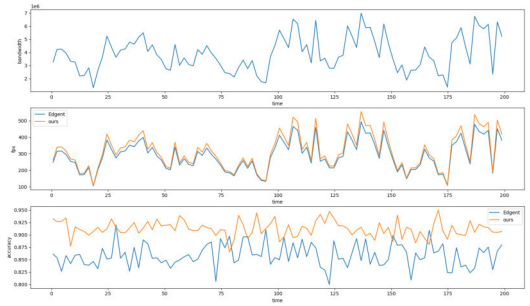


Fig. 1. Graph showing bandwidth, fps and accuracy comparison between our approach and edgent

VI. Conclusion

In this paper we presented our incremental update to the model with early exits. This update represents splits and autoencoder added to the network. Final version of our model can be seen in appendix A. We then implemented simple framework for choosing early exit confidence thresholds and split points similarly to the Edgent framework. Evaluation of our framework shows benefits of using early exits and splits, while showing improvements over Edgent framework. In our future work we want to bring this model into larger scale by having multiple users and multiple servers. We also want to change our $exit\_regressor$ to a deep learning model. This deep learning model will also use V2V and/or V2I to make agents cooperate with early exit and split point decisions

[1] https://users.ugent.be/~jvdrhoof/dataset-4g/

TABLE I
NUMERICAL ANALYSIS OF AUTOENCODER COMPRESSION

1st Split (16x16x64)				2nd Split (8x8x128)				3rd Split (4x4x256)				4th Split (2x2x512)				5th Split (2x2x512)			
Ch	SS	Acc	b	Ch	SS	Acc	b	Ch	SS	Acc	b	Ch	SS	Acc	b	Ch	SS	Acc	b
2	96.9%	0.521	2048	4	96.9%	0.655	1024	8	96.9%	0.776	512	4	99.2%	0.766	64	2	99.6%	0.706	32
4	93.8%	0.733	4096	8	93.8%	0.818	2048	16	93.8%	0.823	1024	8	98.4%	0.843	128	3	99.4%	0.826	48
6	90.6%	0.789	6144	12	90.6%	0.836	3072	24	90.6%	0.835	1536	12	97.7%	0.844	192	4	99.2%	0.849	64
8	87.5%	0.835	8192	16	87.5%	0.843	4096	32	87.5%	0.838	2048	16	96.9%	0.853	256	5	99%	0.854	80
10	**84.4%**	**0.846**	**10240**	**20**	**84.4%**	**0.845**	**5120**	**40**	**84.4%**	**0.844**	**2560**	20	96.1%	0.851	320	**6**	**98.8%**	**0.855**	**96**
12	81.3%	0.849	12288	24	81.3%	0.848	6144	48	81.3%	0.848	3072	**24**	**95.3%**	**0.857**	**384**	7	98.6%	0.854	112
14	78.1%	0.85	14336	28	78.1%	0.848	7168	56	78.1%	0.848	3584	28	94.5%	0.855	448	8	98.4%	0.854	128
16	75%	0.851	16384	32	75%	0.85	8192	64	75%	0.847	4096	32	93.8%	0.857	512	9	98.2%	0.855	144
18	71.9%	0.851	18432	36	71.9%	0.852	9216	72	71.9%	0.85	4608	36	93%	0.858	576	10	98%	0.854	160
20	68.8%	0.853	20480	40	68.8%	0.853	10240	88	65.6%	0.851	5632	40	92.2%	0.858	640	11	97.9%	0.856	176
22	65.6%	0.854	22528	44	65.6%	0.854	11264	96	62.5%	0.853	6144	44	91.4%	0.855	704	12	97.7%	0.855	192
24	62.5%	0.852	24576	48	62.5%	0.854	12288	104	59.4%	0.853	6656	48	90.6%	0.859	768	13	97.5%	0.855	208

Accuracy of the model without splitting was 0.855
Ch - number of channels for compression
SS - Space Saving
Acc - accuracy
b - size of the tensor being sent in bytes

APPENDIX A
IMAGE OF OUR FINAL MODEL

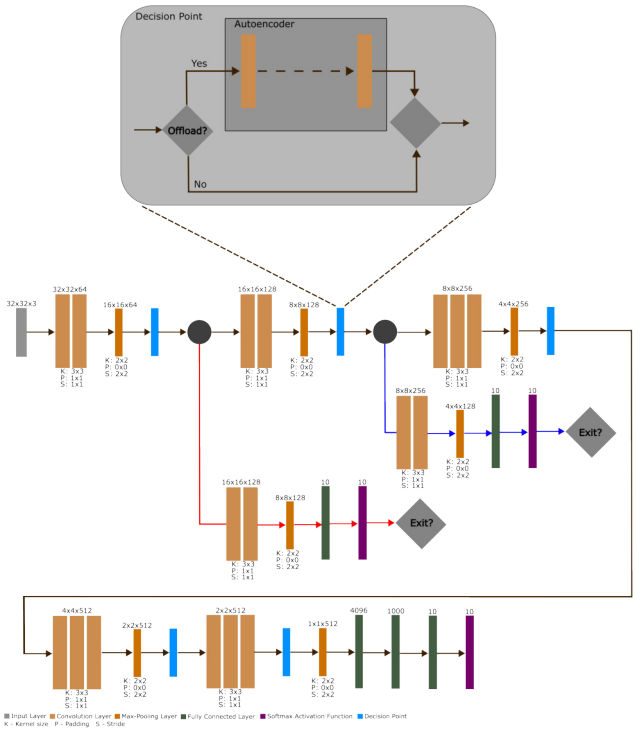


Fig. 2. Final VGG-16 model with early exits and splits

ACKNOWLEDGMENT

This work was supported by The Slovak Research and Development Agency project no. APVV-18-0214 and no. APVV SK-CZ-RD-21-0028.

REFERENCES

[1] D. C. Nguyen, M. Ding, P. N. Pathirana, A. Seneviratne, J. Li, D. Niyato, O. Dobre, and H. V. Poor, "6g internet of things: A comprehensive survey," *IEEE Internet of Things Journal*, vol. 9, no. 1, pp. 359–383, 2021.

[2] Y. C. Hu, M. Patel, D. Sabella, N. Sprecher, and V. Young, "Mobile edge computingâĂŤa key technology towards 5g," *ETSI white paper*, vol. 11, no. 11, pp. 1–16, 2015.

[3] S. Raza, S. Wang, M. Ahmed, M. R. Anwar *et al.*, "A survey on vehicular edge computing: architecture, applications, technical issues, and future directions," *Wireless Communications and Mobile Computing*, vol. 2019, 2019.

[4] E. Li, L. Zeng, Z. Zhou, and X. Chen, "Edge ai: On-demand accelerating deep neural network inference via edge computing," *IEEE Transactions on Wireless Communications*, vol. 19, no. 1, pp. 447–457, 2019.

[5] S. Teerapittayanon, B. McDanel, and H. Kung, "Branchynet: Fast inference via early exiting from deep neural networks," in *2016 23rd International Conference on Pattern Recognition (ICPR)*, 2016, pp. 2464–2469.

[6] S. Laskaridis, A. Kouris, and N. D. Lane, "Adaptive inference through early-exit networks: Design, challenges and directions," in *Proceedings of the 5th International Workshop on Embedded and Mobile Deep Learning*, 2021, pp. 1–6.

[7] Y. Kang, J. Hauswald, C. Gao, A. Rovinski, T. Mudge, J. Mars, and L. Tang, "Neurosurgeon: Collaborative intelligence between the cloud and mobile edge," *ACM SIGARCH Computer Architecture News*, vol. 45, no. 1, pp. 615–629, 2017.

[8] A. E. Eshratifar, A. Esmaili, and M. Pedram, "Bottlenet: A deep learning architecture for intelligent mobile cloud computing services," in *2019 IEEE/ACM International Symposium on Low Power Electronics and Design (ISLPED)*, 2019, pp. 1–6.

[9] R. Rauch, "Task offloading optimization in mobile edge computing architecture," in *22nd Scientific Conference of Young Researchers*, vol. 22, 2022, pp. 104–108.

Robust Data Transmission in 5G Networks Without Infrastructure

[1]*Maros BAUMGARTNER (3rd year)*
Supervisor: [2]Jan PAPAJ

[1,2]Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

[1]maros.baumgartner@tuke.sk, [2]jan.papaj@tuke.sk

***Abstract*— MANET network technologies have undergone major changes since their inception to improve data transmission, routing, security, and robustness. Nowadays, the development of technologies related to MANETs is moving forward rapidly. One such technology is blockchain technology which has the potential to improve these features. This paper offers a brief overview of the proposed routing methods in MANETs using decentralized blockchain technology, artificial intelligence, and game theory in order to achieve robust transmission and routing process in the network.**

***Keywords*—MANET, blockchain, algorithm, robustness, game theory**

I. Introduction

With the advent of fifth- and sixth-generation networks, researchers and research teams are looking to leverage various advanced technologies such as artificial intelligence, machine learning or blockchain technology that are being used to improve network robustness [1]. In our paper, we use decentralized blockchain technology, game theory and artificial intelligence in the routing process to select nodes suitable for routing and data transmission [2].

The present paper shows the possibilities of using decentralized blockchain technology, game theory and artificial intelligence in routing process in MANETs (Mobile Ad-Hoc Network). The paper consists of three chapters. The first chapter focuses on describing and outlining the routing problem in MANETs. The second chapter focuses on a brief description of a novel routing algorithm based on decentralized blockchain technology and game theory to select appropriate nodes for the routing process and subsequent data transmission. The third chapter describes the proposed algorithm used for routing based on decentralized blockchain technology and artificial intelligence to increase the robustness of the network through the routing process.

II. Robust Routing Based on Game Theory and Decentralized Blockchain

A new routing protocol based on decentralized blockchain technology and game theory [3] has been proposed as a new routing method in MANETs. The output of this routing protocol is a found communication route composed of nodes that satisfy the network robustness conditions. Such nodes are considered to be nodes that are robust in terms of reliability. The selection of such nodes is done using Quality of Service (QoS) parameters. These parameters serve as input parameters to cooperative game theory. In this game, communities are formed that consist of at least three nodes in the network. The output of such a game is a node that satisfies the robustness conditions. From such nodes, a robust path is then constructed that is suitable and shows the best results from the data transmission and routing process. Subsequently, such a path is written into the blockchain database. From this database, immutable information about the robustness and reliability of each node in the network is extracted for future routing, which can be used in routing and data transmission to maximize the reliability of the network.

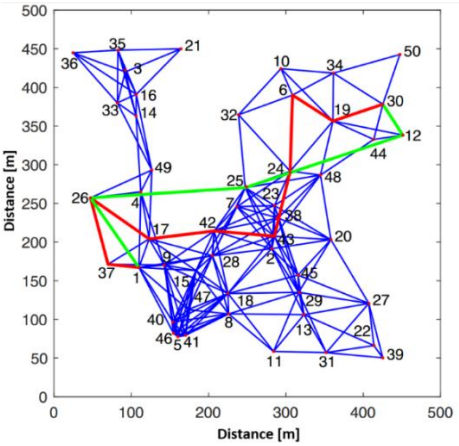


Figure 1 Example of selected routing path based on new routing algorithm

The Figure 1 shows a network with 50 mobile nodes of MANET network where the selected path is highlighted in green color using Dijkastra algorithm [4] which is used to find the shortest path from source node to destination node. The path found through our proposed routing protocol based on game theory and decentralized blockchain technology is highlighted in red. As a result, we compare the communication path length through Dijkastra algorithm and the new routing protocol, where the path found through the routing protocol is longer in terms of number of hops but more reliable in terms of robustness, QoS parameters.

III. New Robust Routing Model for MANET

A novel routing approach based on decentralized blockchain technology and artificial intelligence (AI) is presented in this paper. The proposed algorithm has been implemented in Ad-Hoc On-Demand Distance Vector (AODV) [5], Optimized Link State Routing (OLSR) [6] routing protocols.

The high-level architecture of the proposed system is shown

in the Figure 2. One of the main elements of the algorithm is the computation of routing parameters, these serve as input parameters to the AI, based on which a node suitable for routing is determined, these parameters are also fed into the blockchain [7] where they are subsequently verified using the consensus algorithm. Based on the output data from the AI, a node suitable for communication or routing is determined in terms of performance and reliability. Based on this data, the routing process is carried out. A communication route is selected that meets the conditions for robust data transmission.

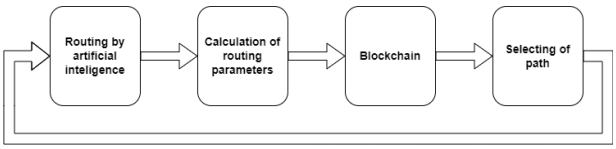


Figure 2 Block diagram of routing algorithm

The actual learning and updating of the network from the perspective of artificial intelligence and decentralized blockchain technology is described at the Figure 3.

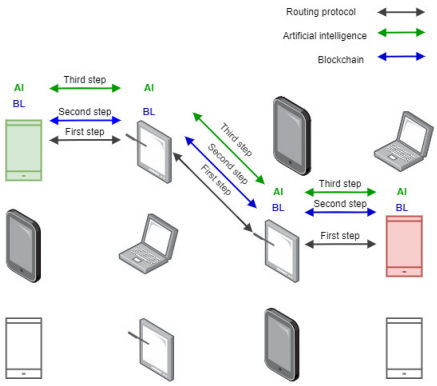


Figure 3 Example of routing based on artificial intelligence and decentralized blockchain technology

Where the first step is to find the communication route through the routing protocol. This is then followed by data transmission and calculation of input technical and network parameters, which are then input into the AI. After successful routing and data transmission, the data obtained is written into the blockchain database. This database serves as an immutable repository of information, which will be used for re-routing in the network to increase the robustness of the network.

The Figure 4 represents a network without infrastructure, in which the decentralized blockchain technology and artificial intelligence have been implemented. There are twelve nodes in the given network, of which one is a source node (green) and one is a destination node (red). Multiple paths have been found in the multi-path routing framework. However, from the point of view of robust data transmission, not all paths or nodes meet the specified conditions. Paths that were found but did not meet the conditions for robust transmission in terms of performance, reliability or security are highlighted in grey. Paths that meet the conditions and can be used for data transmission are highlighted in green. From the simulated network, it can be observed that the proposed algorithm does not search for routing paths based on their length or number of hops only, as classical routing algorithms do, but selects individual nodes based on their parameters as described above. Thus, when comparing the communication route found by the routing protocol and the proposed routing algorithm, we can see that the route found by the routing protocol has fewer hops but does not satisfy the conditions of robust data transmission compared to the route found by the proposed routing protocol.

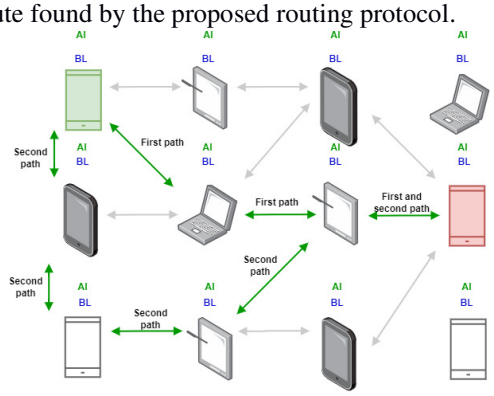


Figure 4 Simulated network with routing based on decentralized blockchain technology and artificial intelligence

IV. Conclusion

The paper demonstrates new ways of routing in MANETs using decentralized blockchain technology, AI and game theory. Based on these technologies, routing algorithms have been proposed to increase the robustness of the network. Future research will be extended to implement different types of artificial intelligence or machine learning and experimentally test them with specific routing protocols used in MANETs in fifth and sixth generation environments.

Acknowledgment

This research was funded by the Slovak Research and Development Agency, research grant no. APVV-17-0208 and VEGA 1/0260/23.

References

[1] Bertin E., Crespi N., Magedanz T., et al: Shaping Future 6G Networks: Needs, Impacts, and Technologies. John Wiley & Sons, 2021.

[2] Chen H., et al: 6G Wireless Communications: Security Technologies and Research Challenges. In: 2020 International Conference on Urban Engineering and Management Science (ICUEMS). IEEE 2020. p. 592-595.

[3] L. L. Njilla, H. N. Ouete and D. K. Doungwa, "Monitoring colluding behavior in MANETs using game theory," 2016 IEEE 21st International Workshop on Computer Aided Modelling and Design of Communication Links and Networks (CAMAD), Toronto, ON, Canada, 2016, pp. 152-153, doi: 10.1109/CAMAD.2016.7790348.

[4] P. E. Abi-Char and A. Youssef, "A Probability-Based Approach for Solving Shortest Path Problems in Gaussian Networks," 2019 IEEE Wireless Communications and Networking Conference (WCNC), Marrakesh, Morocco, 2019, pp. 1-6, doi:10.1109/WCNC.2019.8886027.

[5] M. Singh and J. Sharma, "Performance analysis of secure & efficient AODV (SE-AODV) with AODV routing protocol using NS2," *Proceedings of 3rd International Conference on Reliability, Infocom Technologies and Optimization*, Noida, India, 2014, pp. 1-6, doi: 10.1109/ICRITO.2014.7014729.

[6] Z. Yan, S. Zhai and W. Wang, "Performance Analysis and Improved of Using OLSR in Low-orbit Satellite Communication System," 2022 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA), Dalian, China, 2022, pp. 1423-1426, doi: 10.1109/AEECA55500.2022.9918992.

[7] D. Li, Q. Guo, D. Bai and W. Zhang, "Research and Implementation on the Operation and Transaction System Based on Blockchain Technology for Virtual Power Plant," 2022 International Conference on Blockchain Technology and Information Security (ICBCTIS), Huaihua City, China, 2022, pp. 165-170, doi: 10.1109/ICBCTIS55569.2022.00046.

Overview of Energy Trilemma for Commodity Trading

[1,2]*Nikola HRABOVSKÁ (1st year),*
Supervisor: [2]*Ivana BUDINSKÁ, Consultants:* [1]*Iveta ZOLOTOVÁ,*[1]*Erik KAJÁTI*

[1]Dept. of Cybernetics and Artificial Intelligence, FEEI TU of Košice, Slovak Republic
[2]Institute of Informatics, Slovak Academy of Sciences, Bratislava, Slovak Republic

nikola.hrabovska@tuke.sk, ivana.budinska@savba.sk, iveta.zolotova@tuke.sk, erik.kajati@tuke.sk

***Abstract*—Recently the European and global natural gas markets were hit by a major change in the form of restrictions on gas supplies by Russia. Rough conditions and volatile prices caused complications for gas-dependent households, industries, and many gas-importing countries worldwide. Within the framework of the current global energy crisis, the development of the European electricity and natural gas markets is examined, the development of liquefied natural gas (LNG) in the world markets is examined, and its impact on the gas balance of Europe is evaluated. This paper serves as a brief introduction to the field of research on this issue. The motivation is to work with commodity data and propose a framework to adapt effectively to new conditions.**

***Index Terms*—Demand Prediction, Energy Trilemma, Industry 5.0, Resilience, Sustainability**

I. Introduction

Over the last decade, the energy trilemma has been a prominent pillar of energy policy thinking, concentrating on the balance between attaining sustainability (decarbonization of the energy system), equity (accessibility and affordability for consumers), and energy security (ensuring adequate supply) [1]. Growing concern about the environmental impact of the energy economy in developed economies meant that sustainability had increasingly become a priority, particularly since the Paris Agreement at COP21 in 2015 [2], with affordability and accessibility being related concerns of the energy transition, while energy security was seen as a lesser issue due to the availability of diverse supplies delivered through increasingly fungible global markets. In contrast, in many parts of the developing world, energy fairness and security are more important than sustainability, even though the increasing frequency and severity of extreme weather events have pushed the low-carbon transition to the top of policy priorities. The COVID-19 epidemic and the global recession in 2020 have also intensified the situation from 2020 to 2021 and for many years to come [3].

Russia's invasion of Ukraine has not only forced a significant rethinking of this priority, with energy security being the number one priority internationally, but it has also revealed the disparities in global perceptions of the trilemma. Governments' actions emphasize the disparities in perception and capacity to address the energy trilemma. Finance is also a key point of contention, particularly in the Global South, where countries are requesting assistance from rich countries, which are blamed for the present environmental problems. Further debate over the global rebalancing of the energy trilemma will take place in 2023, with several significant issues poised to dominate policymakers, business decision-making, and academic debate.

By literally powering the world and feeding our planet, commodities are critical to the global economy. Based on the various external factors that significantly influence the commodities sector and the scientific studies that describe the volatility in the commodities market, we see great potential to explore this sector and contribute new knowledge that can push the boundaries of possible solutions to this issue in the commodities market [4], [5], [6]. In this study, we will introduce the term Industry 5.0, focusing specifically on the pillar of resiliency. Resiliency has become one of the commodities sector's top priorities, which has avoided publicity for many years.

II. Energy Trilemma

The energy trilemma is the search for a balance between the three pillars of security, affordability, and sustainability (Fig. 1). It is an important concept that helps us to highlight three distinct areas within the energy sector. It provides a different perspective for the companies involved by pointing out the three dimensions on which to focus rather than just one aspect.

Energy security refers to the ability of countries to secure their energy needs without the risk of supply disruptions. Energy security includes issues relating to political stability in countries that are major energy suppliers, as well as the capacity of countries to diversify their energy supplies.

Energy sustainability focuses on protecting the environment and improving energy efficiency and renewable energy sources. This factor also relates to the long-term impacts of fossil fuel use on climate change and the need to switch to sustainable energy sources.

Affordability refers to the ability of countries and individuals to access energy at reasonable prices. High energy prices

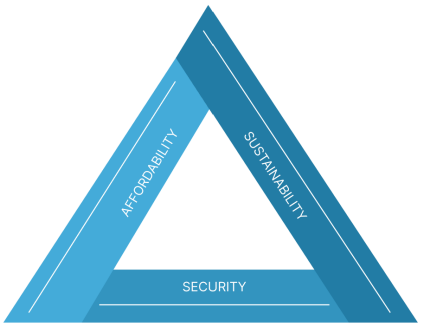


Fig. 1. Energy Trilemma

can have a negative impact on economic growth and cause energy poverty.

The Global Energy Trilemma highlights the importance for countries to balance these three factors. For example, countries seeking to increase energy security may face higher costs in diversifying energy sources. At the same time, countries seeking to achieve energy sustainability may face constraints concerning the availability of specific energy sources.

Scientific studies related to the energy trilemma, and commodity research in general, provide a necessary overview of the field, which can provide a valuable basis for future projects and thus contribute to the global issues in the sector.

III. Industry 5.0

The fourth industrial revolution, Industry 4.0, connects the virtual and physical worlds in the industry to achieve operational productivity and efficiency. The necessity for a new industrial paradigm beyond Industry 4.0 has grown over time in response to critical and complicated societal and economic concerns like:

- Climate change and the collapse of biodiversity.
- Resource scarcity.
- Global shocks such as the COVID-19 pandemic and, more recently, the war in Ukraine.

Purely industrial methods cannot effectively solve those problems. By implementing new digital technologies, new materials, and new processes, for instance, the current Industry 4.0 standard seeks to redefine the industrial sector. A group of specialists [7] contends that Industry 4.0 is fundamentally a technology-centered and growth-oriented industrial paradigm in the recent policy paper Industry 5.0, a transformative vision for Europe - Governing systemic transformations towards a sustainable industry. It fails to adequately consider the economic activity's impact on the environment, society, and sustainable development. Industry 5.0 aims to offer such a response and "provides a vision of the industry that aims beyond efficiency and productivity as the sole goals and reinforces the role and the contribution of industry to society" [8]. By particularly placing research and innovation at the service of the transition to a sustainable, human-centered, and resilient European industry, it does not replace but rather enhances the current "Industry 4.0" strategy.

The following technologies underlying the idea of Industry 5.0 [9]:

- Technologies that connect and integrate the strengths of humans and machines, including human-centric solutions.
- Smart materials with embedded sensors and improved functionalities are made possible by bio-inspired technology and are recyclable.
- Real-time-based digital twins and simulation to model entire systems.
- Technology for secure data transfer, storage, and analysis that can handle data and systems' interoperability.
- Artificial intelligence, for instance, identifies causalities in complex, dynamic systems, leading to actionable intelligence.
- Technologies for reliable autonomy and energy efficiency, such as the ones mentioned above, will demand a lot of energy.

Several issues that can be resolved with appropriate enablers must be taken into consideration for a systematic approach [9]:

- Measures in the governmental and political spheres are necessary due to the rapid pace of transition. This entails employing "agile government" techniques and comprehending the intricate webs of industrial ecosystems and labor marketplaces.
- In the economic component, it is necessary to find ways to retain economic competitiveness and profitability as well as the necessary funding, such as through creating appropriate business models that value ecological and social factors.

It is possible that the energy trilemma and Industry 5.0 will overlap, as both concepts aim to secure a stable and sustainable future. Industry 5.0 can help address energy challenges by increasing energy production efficiency and using renewable energy sources. Digitalization and automation can enable smarter management and monitoring of energy systems and ensure their reliable operation.

As outlined in the Fig. 2, Industry 5.0 has three key pillars that significantly impact corporate strategy: human-centricity, resiliency, and sustainability. However, this article will focus only on the pillars of sustainability and resilience.

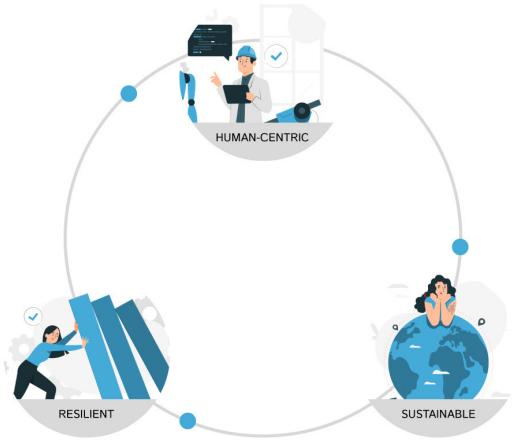


Fig. 2. Three Pillars of Industry 5.0

A. Sustainability

Sustainability refers to protecting the environment, conserving resources, and ensuring the sustainability of society. Today's Industry 4.0 focuses on improving the performance and efficiency of industrial processes but often fails to consider environmental and social issues. It is, therefore, important to consider sustainability in Industry 5.0. This can mean using renewable energy sources, recycling waste, and minimizing the industry's environmental impact. In addition, the use of Industry 5.0 can enable more efficient use of resources, improving sustainability. Overall, it is crucial to consider sustainability issues in the design and implementation of Industry 5.0 to achieve a sustainable and efficient industry. Industry 5.0 is transforming production systems worldwide by removing repetitive jobs from human workers [10], [11].

B. Resilience

Resilience means the ability to withstand various challenges, such as natural disasters, economic crises, or pandemics. In the industry, resilience means adapting and reacting quickly to environmental changes to minimize the consequences. Given current global challenges such as pandemics or climate change, Industry 5.0 and resilience must go hand in hand. This means investing in industrial modernization at the same time as developing strategies to improve resilience. In this way, Industry 5.0 can be more efficient and competitive and better adapted to cope with today's challenges [12], [13].

C. Big data analytics

Industry 5.0 permits the use of 3D symmetry in the creation of innovation ecosystems. Matheus et al. [14] mentioned that big data analytics is a complex procedure for examining big data to uncover data such as hidden patterns and market trends.

It employs a sophisticated analytic approach with a wide range of data sets, including structured and semi-structured data. It has enormous data sets that must be stored and processed using standard tools [15]. It is utilized as real-time data to improve the company industry's competitive advantages, with an emphasis on delivering possible suggestions on predictive discoveries. Most firms rely on big data analytics to make strategic choices [16].

D. Challenges of industry 5.0

Adopting modern technology necessitates more excellent time and effort for human workers. For industry 5.0, customized software-connected factories, collaborative robots, artificial intelligence, real-time information, and the internet of things must be implemented [17].

Establishing trust in ecosystems is vital; therefore, security is a challenge for Industry 5.0. The scale of authentication employed in the industry is the scale to interact with various devices to combat future quantum computing applications to install Internet of Things (IoT) nodes [18], [19]. Artificial intelligence and automation in industry 5.0 is a threat to the business. Thus it must have trustworthy security [20].

IV. Europe Gas Demand and Storage

Concerns about energy security and new policies lead to the largest-ever upward revision of the International Energy Agency's (IEA's) renewable power forecast [21]. Russia's invasion of Ukraine has provided renewables with unprecedented momentum. Significant changes in gas supply sources and flow patterns have occurred in the last 3 to 4 years as new infrastructure has come online and regulatory changes to allow reverse flows have been implemented. South-East Europe was no longer referred to as "Russia's market" even before February 2022. It had demonstrated natural supply diversity in the previous 2 to 3 years, importing and distributing regasified LNG and Caspian (Southern Gas Corridor) gas. Connectivity with other suppliers and interconnectivity within the region has dramatically improved and will continue to improve as additional projects are developed over the next 4 to 5 years. According to Al Jazeera's statistics examination from BP's Statistical Review of World Energy, gas provided around one-third of Europe's energy last year. According to the European Commission, Russia supplied nearly 40% of all imported gas to the EU before the invasion [22].

By having a milder winter in 2022, the healthy level of storage was largely impacted as gas demand decreased. The more substantial non-Russian gas imports and reduced consumption enabled Europe to overcome Russia's gas supply restrictions and achieve near-record storage levels. In 2022, storage injections were 22%, or 13 billion cubic meters (bcm), more than the 5-year average [23]. At the start of November, EU storage facilities were over 95% filled, well above the European Union's 80% objective and in line with the IEA's 10-Point Plan to Reduce the European Union's Reliance on Russian Natural Gas [23].

Currently, European gas prices are showing considerable volatility in the market. This is due to several factors, including increased demand for gas, limited production from large gas fields, low temperatures, and supply problems from some countries. Prices had already risen in 2021 as the world recovered from COVID-19 and supplies became increasingly constrained. As soon as Russia invaded Ukraine, prices skyrocketed. They then began to decline as Russian flows remained high. Prices began to rise in mid-June as flows along Nord Stream began to fall, peaking at the end of August as Nord Stream was finally closed completely. Prices began to fall gradually as LNG imports into Europe remained high, and storage became full. With Russian flows now at deficient levels, the risk premium associated with these flows disappearing is much lower, as evidenced by falling prices. Prices began to rise again in November as the weather worsened and storage withdrawal increased.

Russia's pipeline natural gas deliveries to the European Union were cut in half in the first ten months of 2022, a drop of 60 bcm from the same period in 2021. For the entire year, Russia's pipeline supplies are expected to fall by more than 55%, or 80 bcm, putting unprecedented pressure on both European and global gas markets [24].

V. Future Work

In our research, we will work with commodity data, where we would like to propose a framework to adapt effectively to new conditions and keep production running. We are also establishing an EP Innovation Center (EPIC) in the UVP Technicom building on the TUKE campus to build international relations concerning this research topic and to spread awareness among researchers. We see great potential to pursue further research and analyze the potential applications of Industry 5.0 concepts in energy markets. We can relate the concept of Industry 5.0 to our work, as we have already described that two of the three pillars, resiliency, and sustainability, are essential aspects of this paper's framework.

Many technologies and methods are used to address energy challenges. Using big data analytics, artificial intelligence, and machine learning, we will explore the following areas:

- Renewable energy sources, such as solar, wind, hydro, and geothermal, are increasingly popular and more sustainable than fossil fuel sources. These energy sources help reduce greenhouse gas emissions and dependence on fossil fuels.
- Improving energy efficiency: improving energy efficiency is another effective method of addressing energy challenges. It reduces energy consumption and, thus, energy costs.
- Energy storage: energy storage can help harness energy from renewable sources when unavailable. Energy storage also allows for better planning and management of energy distribution.
- Nuclear power: Nuclear power is an energy source that can provide large amounts of energy with low greenhouse gas emissions. However, nuclear energy has risks and concerns, including safety and waste management.
- Smart grids: smart grids improve traditional energy transmission and distribution networks. They help to optimize energy distribution and enable better integration of renewable energy sources.

These technologies and methods can be used together or separately to address energy challenges. Each of these technologies and methods has advantages and disadvantages, which we intend to explore in our following papers and provide the most appropriate solutions for the general public given the specific situation.

Acknowledgment

This work was supported by the Ministry of Education, Science, Research and Sport of the Slovak Republic and the Slovak Academy of Sciences VEGA 2/0135/23.

References

[1] L. M. Grigoryev and D. D. Medzhidova, "Global energy trilemma," *Russian Journal of Economics*, vol. 6, no. 4, pp. 437–462, 2020.

[2] UNFCCC, "Report of the Conference of the Parties on its twenty-first session, held in Paris from 30 November to 13 December 2015," *United Nations*, Jan. 2016. [Online]. Available: https://unfccc.int/sites/default/files/resource/docs/2015/cop21/eng/10a01.pdf

[3] S. Bobylev and L. Grigoryev, "In search of the contours of the post-covid sustainable development goals: The case of brics," *BRICS Journal of Economics*, vol. 1, no. 2, pp. 4–24, 2020.

[4] D. Bakas and A. Triantafyllou, "Commodity price volatility and the economic uncertainty of pandemics," *Economics Letters*, vol. 193, p. 109283, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165176520301890

[5] Z. Umar, M. Gubareva, and T. Teplova, "The impact of covid-19 on commodity markets volatility: Analyzing time-frequency relations between commodity prices and coronavirus panic levels," *Resources Policy*, vol. 73, p. 102164, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0301420721001781

[6] D. Bakas and A. Triantafyllou, "Volatility forecasting in commodity markets using macro uncertainty," *Energy Economics*, vol. 81, pp. 79–94, 2019.

[7] S. Dixson-Decleve, P. Balland, F. Bria, K. Dunlop, E. Giovannini, D. Tataj, C. Hidalgo, A. Huang, D. Isaksson, F. Martins *et al.*, "Industry 5.0: A transformative vision for europe," 2022.

[8] "Industry 5.0. research and innovation." 2023. [Online]. Available: https://research-and-innovation.ec.europa.eu/research-area/industrial-research-and-innovation/industry-50_en

[9] J. Müller, "Enabling technologies for industry 5.0," *European Commission*, pp. 8–10, 2020.

[10] A. B. Alhassan, X. Zhang, H. Shen, and H. Xu, "Power transmission line inspection robots: A review, trends and challenges for future research," *International Journal of Electrical Power & Energy Systems*, vol. 118, p. 105862, 2020.

[11] F. Aslam, W. Aimin, M. Li, and K. Ur Rehman, "Innovation in the era of iot and industry 5.0: Absolute innovation management (aim) framework," *Information*, vol. 11, no. 2, p. 124, 2020.

[12] X. Xu, Y. Lu, B. Vogel-Heuser, and L. Wang, "Industry 4.0 and industry 5.0—inception, conception and perception," *Journal of Manufacturing Systems*, vol. 61, pp. 530–535, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0278612521002119

[13] M. Breque, L. De Nul, and A. Petridis, "Industry 5.0: towards a sustainable, human-centric and resilient european industry," *Luxembourg, LU: European Commission, Directorate-General for Research and Innovation*, 2021.

[14] L. E. M. Matheus, A. B. Vieira, L. F. Vieira, M. A. Vieira, and O. Gnawali, "Visible light communication: concepts, applications and challenges," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 4, pp. 3204–3237, 2019.

[15] J. L. Hopkins, "An investigation into emerging industry 4.0 technologies as drivers of supply chain innovation in australia," *Computers in Industry*, vol. 125, p. 103323, 2021.

[16] A. Kumar, A. Choudhary, H. Kaur, S. Guha, S. Mehta, and A. Husen, "Potential applications of engineered nanoparticles in plant disease management: a critical update," *Chemosphere*, vol. 295, p. 133798, 2022.

[17] H. P. Tripathy and P. Pattanaik, "Birth of industry 5.0:"the internet of things" and next-generation technology policy," *Int J Adv Res Eng Technol*, vol. 11, no. 11, pp. 1904–1910, 2020.

[18] M. Ghobakhloo, M. Fathi, M. Iranmanesh, P. Maroufkhani, and M. E. Morales, "Industry 4.0 ten years on: A bibliometric and systematic review of concepts, sustainability value drivers, and success determinants," *Journal of Cleaner Production*, vol. 302, p. 127052, 2021.

[19] M. Doyle Kent and P. Kopacek, "Do we need synchronization of the human and robotics to make industry 5.0 a success story?" in *Digital Conversion on the Way to Industry 4.0: Selected Papers from ISPR2020, September 24-26, 2020 Online-Turkey*. Springer, 2021, pp. 302–311.

[20] E. Tijan, S. Aksentijević, K. Ivanić, and M. Jardas, "Blockchain technology implementation in logistics," *Sustainability*, vol. 11, no. 4, p. 1185, 2019.

[21] IEA, "Executive summary – renewables 2022 – analysis." [Online]. Available: https://www.iea.org/reports/renewables-2022/executive-summary

[22] J. Korosteleva, "The implications of russia's invasion of ukraine for the eu energy market and businesses," *British Journal of Management*, vol. 33, no. 4, pp. 1678–1682, 2022.

[23] IEA, "Never too early to prepare for next winter–analysis." [Online]. Available: https://www.iea.org/reports/never-too-early-to-prepare-for-next-winter

[24] iea, "Executive summary – gas market report, q2-2022 – analysis." [Online]. Available: https://www.iea.org/reports/gas-market-report-q2-2022/executive-summary

Method of data hiding represented by QR codes in images by using XOR logical operation

[1]*Samuel ANDREJČÍK (2nd year)*
Supervisor: [2]Ľuboš OVSENÍK

[1,2]Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

[1]samuel.andrejcik@tuke.sk, [2]lubos.ovsenik@tuke.sk

***Abstract*— This work deals with the digital processing of images through the technique of steganography. Different techniques of steganography solve the problem of data authentication, data security and copyright. Due to the fact that the problem of data security is currently very topical, these techniques are often sought after. This is related to the constant development of new algorithms ensuring solutions that will safely hide the secret message in the image and at the same time will be robust against various types of attacks. The motivation for writing this work and the novel itself lies in the technique of representing a secret message with a QR (Quick Response) code, which is an approach that is not widely explored by the scientific community. The secret message is written in the least significant bit plane of the cover image, which ensures that the modification of the image cannot be detected by the naked eye. The proposed solution is evaluated by mathematical quality assessment parameters PSNR, MSE, NCC and SSIM.**

***Keywords*—Image processing, information hiding, LSB, security, QR code, steganography**

I. INTRODUCTION

Many steganographic algorithms have been proposed so far. They all try to hide information by relying on some of the known techniques [1]. In order to improve the security of steganography, new algorithms have been developed in recent years to hide information by relying on some of these techniques. However, all information hiding techniques have their advantages and disadvantages [2].

In this work, image processing in steganography will be described using the LSB (Least Significant Bit) method, in which information is written by *XOR* operation into the LSB plane (Fig. 1), and then the quality of the proposed solution between the cover image and the stego image is monitored by mathematical evaluations. It is subsequently subjected to cutting and noise attacks, while the degradation of the stego image, carrying a secret message, which is represented by a QR code, is monitored. The suitability of the proposed solution is again compared, but this time between the original secret message and the extracted secret message.

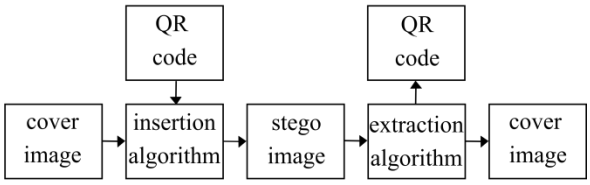


Figure 1 Block diagram showing a steganographic design for hiding data in an image

II. THEORY

Any processing applied to the images may cause a significant loss of information or quality. Image quality assessment methods can be divided into objective and subjective methods [3]. While subjective methods are based on human judgment and operate without additional explicit criteria, objective methods are based on comparisons using explicit numerical criteria.

MSE (Mean Squared Error), PSNR (Peak Signal to Noise Ratio), SSIM (Structural Similarity Index Measure) and NCC (Normalized Cross-Correlation) were used to evaluate the quality of the proposed solution [4].

In order to be able to write the data represented by the QR code into the cover image, it is advisable to use some of the well-known techniques. LSB replacement and LSB matching are among the steganographic methods that are used in the implementation of similar tasks [5].

LSB replacement uses an algorithm that replaces the least significant bits of the cover image with bits of the secret message. It is also one of the simplest steganographic methods.

An extended version is also used in this method, in which several bits of the secret message are written into the last two or more bits of the least significant bit plane, but this extended method is more susceptible to revealing the secret message, as the original image differs significantly more from the stego image. as in the basic method of replacing one, last bit [6].

On the other hand is LSB matching, also known as ±1 embedding. It is an improved LSB replacement method. Both methods are based on hiding secret messages in the LSB plane by embedding, but with a different approach [7].

This approach can be demonstrated as follows: if a secret message with value 1 is inserted into an image byte with intensity $120_{(10)}$, the resulting change in image intensity is $121_{(10)}$. With the ±1 embedding method, 2 possible results are possible, as the intensity value can both increase and decrease

compared to the original intensity. Thus, with an original image intensity of $120_{(10)}$ and a secret message with a value of 1, the resulting image intensity can be $119_{(10)}$, but also $121_{(10)}$, which will create a more robust system that is more resistant to attacks. The resulting intensity of the stego image, i.e. whether the value of the image element will be increased or decreased, is determined by a pseudo-random sequence.

However, when solving this task, the method of binary logical operation XOR (exclusive or) was used, while the impact of the proposed solution on the mathematical evaluation of quality was investigated.

III. Proposed solution

The above-mentioned block diagram (Fig. 1) shows the sequence of the proposed solution. The system serves to cover the very existence of secret communication by modifying the cover image with a secret message. The proposed solution is for images that do not undergo lossy compression. The entire design of the solution, including algorithms and simulation, was implemented in the MATLAB platform, so it is possible to simulate the proposed solution based on the procedures described below.

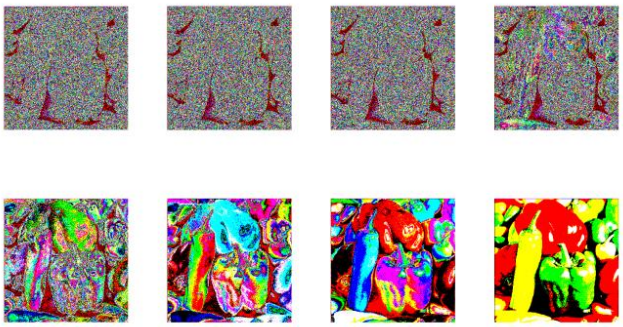


Figure 2 Dividing the image into 8 bit planes

In order for the square QR code to be written evenly in the image and at the same time not a single side overflows through this image, it is necessary to find out which side of the cover image is smaller (this applies to images that are oriented in width or height). Subsequently, it is possible to manually determine how many times the QR code needs to be reduced compared to the cover image, while remembering that the larger the reduction, or loss compression of the QR code, the subsequent correction of the QR code is more difficult.

In order to ensure an even shift of the QR code across the image, it is necessary to find out what the ratio of the QR code and the cover image is.

After the information is successfully written into the LSB plane of the cover image, all the bit planes can be combined again to form a stego image, that is, an image that consists of the original cover image and now the transmitted information.

Table 1 Assessment of the quality of the stego image after inserting a QR code with information about the length of 1000 characters

Image name	PSNR (dB)	SSIM	NCC
baboon.bmp	64.9239	$1\text{-}1.7\cdot10^{-5}$	$1\text{-}3.3\cdot10^{-6}$
baboonG.bmp	60.1526	$1\text{-}1.7\cdot10^{-4}$	$1\text{-}1.7\cdot10^{-5}$
barbaraG.bmp	60.1526	$1\text{-}3.5\cdot10^{-4}$	$1\text{-}4.5\cdot10^{-4}$
cameramanG.bmp	60.1526	$1\text{-}6.5\cdot10^{-4}$	$1\text{-}8.1\cdot10^{-6}$
f16.bmp	64.9239	$1\text{-}1.6\cdot10^{-4}$	$1\text{-}5.4\cdot10^{-6}$
lena.bmp	64.9239	$1\text{-}2.5\cdot10^{-5}$	$1\text{-}3.6\cdot10^{-4}$
lenaG.bmp	60.1527	$1\text{-}5.0\cdot10^{-4}$	$1\text{-}1.7\cdot10^{-5}$
pentagonG.bmp	60.1273	$1\text{-}3.2\cdot10^{-4}$	$1\text{-}4.4\cdot10^{-5}$
peppers.bmp	64.9239	$1\text{-}1.1\cdot10^{-5}$	$1\text{-}3.4\cdot10^{-6}$
sailboat.bmp	64.9239	$1\text{-}2.9\cdot10^{-5}$	$1\text{-}2.3\cdot10^{-6}$

The first experiment was performed within cutting attack on the newly created stego image. This image was subjected to zeroing at all bit levels, resulting in the corners of the stego image being cut off.

Fig. 3 represents the stego image after the cut-out attack and the QR code extracted from this image.

Figure 3 Stego image after the cutting attack and the extracted QR code from the stego image

As can be seen, the percentage damage for the stego image is the same as for the QR code. This is caused by the even distribution of the QR code in the image. If the image were placed in the upper left corner, without adding the uniform displacement functionality, the loss on the decoded code would be much higher.

Table 2 Assessment of the quality of the extracted QR code at 8% cutting from the stego image

Image name	PSNR (dB)	SSIM	NCC
baboon.bmp	14.0974	0.917710	0.92082
baboonG.bmp	14.1129	0.916953	0.92111
barbaraG.bmp	13.8099	0.919813	0.91691
cameramanG.bmp	14.0897	0.918432	0.92071
f16.bmp	14.0475	0.916559	0.91997
lena.bmp	14.0666	0.917995	0.92040
lenaG.bmp	14.1051	0.918102	0.92104
pentagonG.bmp	13.7967	0.913637	0.91528
peppers.bmp	13.9756	0.916153	0.91861
sailboat.bmp	14.1167	0.918428	0.92127
Average	14.0218	0.917378	0.91961

The proposed solution also faced a noise attack separately, using Salt & Pepper noise with a noise density of 0.05 - it affects approximately 5% of all pixels. Noise was applied to whole stego image.

Fig. 4 shows the original and extracted QR code declares that even in the case of a noise attack, the functionality is ensured even with this algorithm.

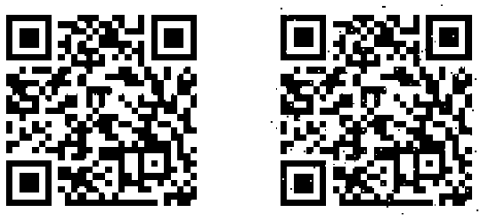


Figure 4 The original QR code and the extracted QR after the noise attack

After reading the information from both of these QR codes, it is possible to declare that the data was not lost and fully resisted the noise attack.

Table 3 Assessment of the quality of the extracted QR code with 5% noise of stego image

Image name	PSNR (dB)	SSIM	NCC
baboon.bmp	16.0356	0.806646	0.94924
baboonG.bmp	15.8757	0.805920	0.94736
barbaraG.bmp	16.0841	0.805704	0.94981
cameramanG.bmp	15.9343	0.802932	0.94805
f16.bmp	15.9876	0.807500	0.94870
lena.bmp	16.2016	0.820067	0.95113
lenaG.bmp	16.1766	0.813924	0.95090
pentagonG.bmp	15.9041	0.585877	0.94772
peppers.bmp	15.9461	0.804019	0.94818
sailboat.bmp	16.0055	0.802430	0.94897
Average	16.0151	0.785502	0.94901

IV. Future Work

To continue this work, it will be necessary to change the QR code encoder, which will not generate a fixed resolution regardless of the length of the transmitted message, and also to change the method of LSB replacement or LSB matching instead of writing the bits of the secret message into the LSB plane of the cover image using the binary logical operation *XOR*.

V. Conclusion

In this work, the main task was to design a new steganographic system that would ensure the hiding of a QR code carrying a secret message of 1000 characters in various images, which are often used in testing steganographic procedures. When implementing the solution, the secret message was hidden in the LSB plane of the cover image.

Cutting and noise attacks were performed on a given stego image. However, after carrying out these attacks, it is clear that the proposed solution is not suitable for security due to the use of an inappropriate QR code encoder, and this encoder have to be changed in the future. The reason is the fact that the coder used generates QR codes with a high resolution, and subsequently it is necessary to implement subsampling, which has a negative impact on the mathematical results. However, thanks to Reed-Solomon coding, the information transmitted by the QR code can be read and is not damaged. By using a more suitable QR code encoder, a significant improvement of the mathematical results, especially the PSNR is expected. The results of the index of structural similarity and the normalized correlation coefficient yielded better results.

Acknowledgment

This research was funded by the Slovak Research and Development Agency, research grant no. APVV-17-0208 and VEGA 1/0260/23.

References

[1] A. Rehman, T. Saba, T. Mahmood, Z. Mehmood, M. Shah, A. Anjum, "Data hiding technique in steganography for information security using number theory," in. *J. of Information Science,* Saudi Arabia, 2018.

[2] T. Mahmood, Z. Mehmood, M. Shah, T. Saba, "A robust technique for copy-move forgery detection and localization in digital images via stationary wavelet and discrete cosine transform." in *J. of Vis. Commun. and Image Represent.*, Pakistan, 2018.

[3] M. R. N. Torkaman, N. S. Kazazi, A. Rouddini, "Innovative Approach to Improve Hybrid Cryptography by Using DNA Steganography," in *Int. J. on New Computer Architectures and Their Applications,* Malaysia, 2012.

[4] J. Fridrich, Steganography in Digital Media. Binghamton, NY: Cambridge University Press, 2014.

[5] A. Pradhan, et al. "Performance evaluation parameters of image steganography techniques," in. *Int. Conf. on Research Advances in Integrated Navigation Systems*, 2016.

[6] J. Fridrich, M. Goljan, R. Du, "Reliable Detection of LSB Steganography in Color and Grayscale Images," in *Proc. of the 2001 Workshop on Multimedia and Security*, 2001.

[7] A. D. Ker, "Steganalysis of LSB matching in grayscale images," in *IEEE Signal Processing Letters*, 2005

Goethite nanorods concentration influence on the E7 liquid crystal properties

[1]*Dmytro Miakota (2nd year)*
Supervisor: [2]Natália TOMAŠOVIČOVÁ

[1,2]Institute of Experimental Physics of the Slovak Academy of Sciences, Košice, Slovak Republic

[1]dmytro.miakota@tuke.sk, [2]nhudak@saske.sk

***Abstract*— Goethite is the most common of the iron oxyhydroxides. The outstanding magnetic properties of its nanorods increase scientist's research interest. The ability of goethite nanorods to align in different directions depending on the strength of the external magnetic field is of particular interest. The submitted work reviews the influence of magnetic and electric field on structure and physical properties of composites based on the goethite nanoparticles and liquid crystal and highlight the significance of their research due to their potential practical application. Within the research, the liquid crystal doped with magnetic nanorods was investigated.**

***Keywords*— liquid crystal, goethite nanoparticles, liquid crystal composites.**

I. Introduction

Goethite is a common iron oxide mineral that was named after the German polymath Johann Wolfgang von Goethe (1749–1832). This mineral is a good conductor of electricity, which makes it useful in the production of electrodes and other electronic components [1].

Goethite usually forms rod-shaped nanoparticles (nanorods), which often lack long-range crystalline order, because the particles consist of many grains, typically around 3-7 nm in size, with low-angle grain boundaries [2].

Aqueous suspensions of goethite nanorods show a self-assembly under application of an external magnetic field that can be semi quantitatively explained in the frame of statistical-physics models of the nematic ordering, i.e. long axis of the nanoparticles are oriented in the same direction denoted by unit vector $\vec{n}$, called director. The main interest of these nematic suspensions rather lies in their original behaviour when submitted to magnetic field. The orientation of long axis of nanoparticles can be perpendicular or parallel to the direction of magnetic field depending on its stress. The reorientation of the director upon increasing magnetic field is due to the negative anisotropy of the magnetic susceptibility. At high field, nematic domains aligned perpendicular to the field and become biaxial [3].

The objective of the research is to experimentally investigate the changes of physical properties of liquid crystal composites doped with goethite nanorods under the influence of magnetic, electric fields and their combinations. To attain the objective, the capacitance measurements were employed as primary experimental method.

II. Magnetic properties of goethite (α-FeOOH) nanorods

In 1902, Majorana examined the magnetic-field-induced birefringence of "fer de Bravais," in a mixture of goethite particles in suspension. He observed field-induced linear birefringence when a transverse magnetic field was applied to this dispersion. This is now known as the Majorana effect [4].

Lemaire et al. studied the properties of pure goethite in dispersion and found that similar birefringence effects were also observed [5]. It was discovered that the board-like goethite particles have a weak permanent magnetic moment along the long axis of the particles.

The model for the magnetic energy of a goethite particle that was mentioned earlier, which is based on the rod-like shape assumption [6], is a simplification, as goethite particles are actually board-like and have three different dimensions. The magnetic energy per particle, E_m, is given by:

$$E_m = -\mu B cos\theta - \frac{\Delta\chi V}{2\mu_0} B^2 cos^2\theta, \qquad (1)$$

where μ is the permanent magnetic moment, $\Delta\chi = \chi_{\parallel} - \chi_{\perp}$ is the magnetic susceptibility anisotropy (which is negative in this case), V is the volume of the particle and θ is the angle between the particle's magnetic moment and the direction of the magnetic field. The first term describes the interaction between the particle's dipole and the field, and the second term describes the induced magnetization due to the magnetic susceptibility anisotropy.

In small fields, the first term of the magnetic energy equation dominates, and the minimum energy is achieved when the particles align parallel to the field (at an angle of $\theta = 0$). However, in large fields, the second term becomes more significant, and the energy minimum is achieved when the particles align perpendicular to the field (at an angle of $\theta = \pi/2$).

Lemaire *et al.* [5] revealed that the goethite particles have a permanent magnetic moment along their long axis, which is attributed to uncompensated spins within their anti-ferromagnetic crystal structure. In contrast, the magnetic easy axis is along the shortest particle dimension. The interplay between these two axes leads to the peculiar reorientation of the particles, from parallel alignment to a small magnetic field to perpendicular alignment to a large magnetic field [7].

III. Liquid Crystals

Liquid crystal materials in the isotropic phase act similarly to any other liquid. However, upon cooling, these materials go through one or more mesophases before reaching the crystalline phase (thermotropic liquid crystals). Mesophases are intermediate states in which the material exhibits properties

of both liquids and solids. Within the mesophase, liquid crystal molecules can be ordered in various ways, such as the nematic phase, in which the elongated molecules of the liquid crystal all point in the same direction. This direction is represented by a director $\vec{n}$ [8].

The response of a liquid crystal to an external electric or magnetic field is typically observed using a cell consisting of two separated glasses with a polymer layer in between. The overall design of the liquid crystal cell is depicted in Fig. 1. In this work the alignment of LC molecules was planar.

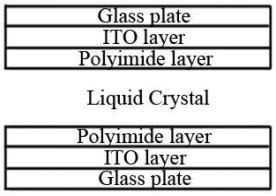


Fig. 1. Liquid crystal cell for optical and dielectric measurements. ITO layer on glass plates serves as electrode and polyimide layer provides required alignment.

When a sufficiently large field, known as the critical or threshold field, is applied to the material, the liquid crystal molecules will begin to reorient. This effect is called the Fréedericksz transition [9].

IV. Materials and Methods

A. E7 Liquid Crystal

E7 is thermotropic, nematic liquid crystal with nematic to isotropic transition temperature at T_{NI} = 61°C and a glass transition temperature at T_g = - 62 °C, providing a large nematic range of approximately 120°C. E7 is widely used because it possess high anisotropy over a large temperature range [10].

B. Samples preparation

The E7 liquid crystal was doped with goethite magnetic nanoparticles (MNPs) at concentrations of 0.1%, 0.05% and 0.01% by volume. The goethite nanorods were synthesized following a recipe in [11]. The microstructural characterization of the nanopowders was performed using a transmission electron microscopy (TEM). The average length, width, and thickness of the goethite nanorods were estimated to be 350 ± 100 nm, 25 ± 7 nm and 10 ± 5 nm respectively.

C. Methods

The transition from the isotropic to the nematic phase was studied using dielectric measurements. The samples, consisting of either undoped E7 or E7 doped with goethite MNPs, were filled into capacitors for the measurement of magnetic and electric Fréedericksz transition and for measurement of the temperature of phase transition from isotropic to nematic phase.

V. Conclusions and Results

Goethite nanorods concentration influence on the E7 liquid crystal properties was explored: on the temperature of the phase transition (isotropic-nematic) (the presence of goethite nanorods reduces the phase transition temperature), on the magnitude of the critical field required for the Fréedericksz transition (electric and magnetic) and on a different combination of external fields. For example, the presence of goethite nanorods reduces the value of the critical field (Fig. 2). All these results point to the excellent interaction of goethite nanoparticles with an external magnetic field. It is also important to note that the influence of the concentration of nanoparticles on the liquid crystal has very little effect on other physical properties of the crystal (such as the phase transition temperature), which may be useful for possible future practical applications.

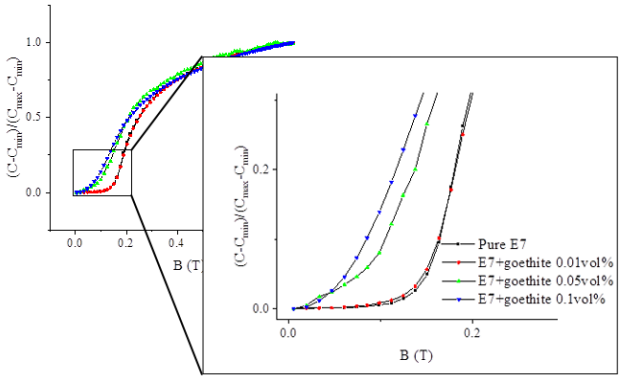


Fig. 2. Magnetic Fréedericksz transition for pure E7 liquid crystal and E7 doped with goethite nanorods of concentration 0.1vol% (blue), 0.05vol% (green) and 0.01vol% (red)

VI. Next Steps

The plans for further research can be classified into two parts: the first part involves the investigation of defects in droplets of liquid crystal composites containing various nanoparticles, including goethite. The second part entails the exploration of the memory effect in liquid crystal composites with nanoparticles, to examine the impact of nanoparticles on the memory effect. Studying the effect of nanoparticle doping on the optical, electrical, and magnetic properties of liquid crystal composites is essential for discovering new properties and effects.

References

[1] "Mineralogy Database: Goethite," [Online]. Available: https://www.mindat.org/min-2077.html.

[2] D. E. Madsen, L. Cervera-Gontard, T. Kasama, R. E. Dunin-Borkowski, C. B. Koch, M. F. Hansen, C. Frandsen and S. Mørup, "Magnetic fluctuations in nanosized goethite (α-FeOOH) grains," *Journal of Physics: Condensed Matter,* vol. 21, no. 1, 2008.

[3] B. J. Lemaire, P. Davidson, D. Petermann, P. Panine, I. Dozov, D. Stoenescu and J. P. Jolivet, "Physical properties of aqueous suspensions of goethite (α-FeOOH) nanorods (Part II: In the nematic phase)," *The European Physical Journal E,* vol. 13, pp. 309-319, 2004.

[4] Q. Majorana, "Sur la biréfringence magnétique," *Compt Rend Acad Sci,* vol. 135, pp. 159-161, 1902.

[5] B. J. Lemaire, P. Davidson, J. Ferré, J. P. Jamet, P. Panine, I. Dozov and J. P. Jolivet, "Outstanding Magnetic Properties of Nematic Suspensions of Goethite (α-FeOOH) Nanorods," *Phys. Rev. Lett.,* vol. 88, no. 12, p. 125507, 2002.

[6] B. J. Lemaire, P. Davidson, J. Ferré, J. P. Jamet, D. Petermann, P. Panine, I. Dozov and J. P. Jolivet, "Physical properties of aqueous suspensions of goethite (α-FeOOH) nanorods (Part I: In the isotropic phase)," *The European Physical Journal E,* vol. 13, pp. 291-308, 2004.

[7] V. d. p. Esther, "Goethite liquid crystals and magnetic field effects," 2010.

[8] J. Svoboda and M. Glogarová, "Kapalné krystaly.," [Online]. Available: http://kubusz.net/Reserse/new/1/9_axial_chirality/9_9_Svoboda.pdf.

[9] I. W. Stewart, "The Static and Dynamic Continuum Theory of Liquid Crystals: A Mathematical Introduction," in *Liquid Crystals Book Series 2*, New York, Taylor & Francis, 2004, p. 360.

[10] A. Bouriche, L. B. Alachaher and U. Maschke, "Phase behaviour and electro-optical response of systems composed of nematic liquid crystals and poly (2-ethylhexylacrylate)," in *Liquid Crystals*, 5 ed., vol. 45, 2018, pp. 656-665.

[11] P. Kopčanský, V. Gdovinová, S. Burylov, N. Burylova, A. Voroshilov, J. Majorošová, F. Agresti, V. Zin, S. Barison, J. Jadzyn and N. Tomašovičová, "The influence of goethite nanorods on structural transitions in liquid," *Journal of Magnetism and Magnetic Materials,* vol. 459, pp. 26-32, 2017.

Detection of Significant Entities on Blockchain Network Using Machine Learning

[1]Tomáš ADAM (2nd year)
Supervisor: [2]František BABIČ

[1,2]Department of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1]tomas.adam@tuke.sk, [2]frantisek.babic@tuke.sk

***Abstract*—This paper presents an anomaly detection approach for preliminary entity identification within the WAX blockchain network using unsupervised machine learning algorithms. The proposed approach was evaluated by utilizing three detection algorithms for assigning an anomaly score to each account in the dataset. The results of this initial analysis indicate the presence of various suspicious activities and highlight the effectiveness of the anomaly score ranking mechanism in identifying such cases.**

***Keywords*—anomaly detection, blockchain, distributed ledger, unsupervised machine learning**

I. Introduction

Nowadays, the popularity of new financial technologies in the field of digital currencies, trading platforms, and payments utilizing blockchain technology observe a significant increase [1]. Blockchain technology offers numerous advantages, such as transparency and decentralization [2], however, distributed ledger payments also bring their own potential risks. One of the main regulatory concerns regarding blockchain technology is its potential use in illegal activities [3], such as wash trading [4], money laundering or terrorism financing [5]. While rule-based approaches [6] are the foundation of decision systems in financial institutes. Machine learning algorithms [7] wider the possibility of detecting suspicious activities and the chance to identify new, previously unknown behavior. However, there are various challenges in the related field with a significant imbalance between legitimate and suspicious behavior. This problem is exacerbated by a lack of general knowledge about anomalous cases, as labeled data are very difficult to obtain.

II. Initial Dataset Preparation

For this initial study, the WAX blockchain was selected for further analysis due to its distinctive features, including short user addresses and the absence of transaction fees. In addition, the lack of transaction fees on the network, encourages users to generate a higher volume of transactions, thereby offering ample opportunities for in-depth analysis [8]. Furthermore, the WAX blockchain's support for the non-fungible token (NFT) standard allows the examination of transactions from diverse consumer perspectives beyond traditional financial behavior.

For the initial dataset preparation, a Hyperion[1] node was established, which provides access to individual blocks of the WAX blockchain network. This node allows data collection to be focused on a specific smart contract, thereby facilitating the tracking of transaction data and standards across the entire network. Given the diversity and large volume of data on the blockchain network, the dataset is being gradually built and currently contains more than 10 million records. Each record includes 13 attributes that characterize specific actions, among which the most interesting are the transaction id, timestamp, sender, receiver and the transferred amount of cryptocurrency.

III. Experiments and Results

A. Data pre-processing

After initial data collection task, the individual transactions were aggregated to improve the definition of accounts during the main analysis. This approach assumed that a suspicious account may exhibit an unusual frequency of transactions with short time intervals between events. The Recency, Frequency, and Monetary (RFM) variables can effectively capture this pattern [7]. Any user who displays an unusual value in one of these variables can be considered suspicious.

The result of computing the RFM variables from the transactions of individual accounts was captured in five-time intervals and represented by the following columns:

- Recency - the time between the last transaction and the end of the time interval.
- Frequency - number of transactions.
- Monetary - value of transactions.

B. Anomaly detection

For conducting the preliminary detection, three algorithms were: utilized Autoencoder (AE) [9], Isolation Forest (IF) [10] and Cluster Based Local Outlier Factor (CBLOF) [11]. These algorithms assigned an anomaly score to each account in the dataset based on the available metrics and various parameters. To ensure comparability between the results of the individual models, all values were rescaled to the interval (0, 1). The primary anomaly score for each account was determined by computing the average of the scores assigned by the three previous algorithms. This score represents a useful ranking mechanism, with higher values indicating a greater deviation from the normal behavior. The overall distribution of these scores is depicted in Fig. 1, where the color scale represents the average evaluation of the selected models.

Most of the individual scores are mostly aligned along the diagonal of the graph, demonstrating a consistent evaluation of the cases by the models. The primary anomaly score orders the records into a list, with the accounts at the top indicating the most suspicious cases that require further manual analysis. This facilitates verification by human experts and a deeper analysis of user behavior, which can be used to label the transactions in the dataset and classify future cases.

[1] Hyperion History API: https://github.com/eosrio/hyperion-history-api

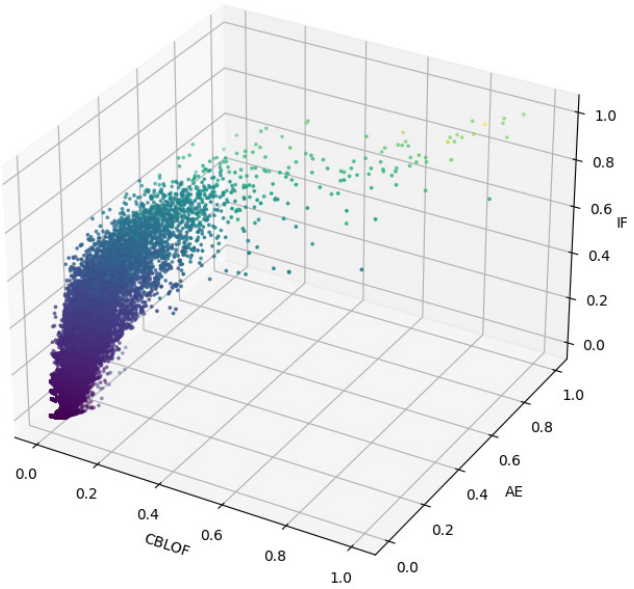


Fig. 1. Visualization of the average anomaly score

C. Results

The initial analysis of the highest ranked accounts revealed several entities that are responsible for a large percentage of transactions on the network. Many of these accounts represent connections (gateways) to the blockchain network as known, centralized and decentralized exchanges along with digital asset marketplaces. The decentralized public architecture of these entities granted access to the blacklists of these smart contracts, leading to the extraction of fraudulent accounts with restricted access to the platforms.

Moreover, several records were identified that represent instances of automated arbitrage, with a significantly higher number of transactions compared to normal accounts. While arbitrage is not illegal, it shares a common behavior pattern with the wash trade techniques, where the involved accounts engage in an excessively large number of buy and sell orders.

Additionally, the analysis revealed the presence of several accounts with an excessively high number of orders created and processed on an NFT digital marketplace. Conduction of an in-depth investigation revealed that these accounts had been reported multiple times by the users and they are linked to a widely known phishing scam network. As of 2023, this organized fraudulent network is still active, and the estimated value of stolen digital funds in verified cases exceeds one hundred thousand dollars. However, it should be noted that due to the volatility of the cryptocurrency market, the nominal value can be even higher, potentially reaching up to ten times the documented amount.

Manual analysis of this results is time-consuming, and the algorithms used to detect anomalies do not offer a clear interpretation of their decisions. Nevertheless, the WAX community is very active, and its members can be considered experts in the related field. Discussions regarding unusual transactions and suspicious accounts can be found in the official fraud reporting channels on community forums, which can be further used for labeling.

IV. Conclusion and Future Work

Implementation of the unsupervised detection method and analysis of social media messages led to the identification of over 60 entities, cryptocurrency exchanges, and decentralized applications. Along with the detection of expected anomalies such as arbitrage bots and potential wash trading, the analysis also uncovered previously unknown cases of phishing attacks and fraudulent transactions. After a comprehensive analysis of community forums and smart contracts of the most important entities on the blockchain network, over 250 accounts were identified to participate in various fraudulent schemes.

Transactions made by these accounts will serve as a focal point for future research tasks, with detection efforts aimed at verifying and classifying suspicious transactions, as well as monitoring users' behavior from market abuse and anti-money laundering perspective. Since the simple interpretation of the outcome is a requirement for systems used in the traditional financial sector. This limitation presents an opportunity for exploring the potential of Explainable Artificial Intelligence models in future research increasing the overall transparency of the anomaly detection process.

The analysis of the WAX blockchain network represents the initial phase in acquiring data and deeper information to facilitate the further implementation of anomaly detection. The use of supervised learning algorithms in future research can offer better accuracy in identifying anomalous behavior by leveraging labeled data and accumulated knowledge.

Acknowledgment

The work was supported by The Slovak Research and Development Agency under grant no. APVV-20-0232 and The Scientific Grant Agency of the Ministry of Education, Science, Research and Sport of the Slovak Republic under grant no. VEGA 1/0685/21.

References

[1] Morkunas, V. J., Paschen, J., & Boon, E. (2019). How blockchain technologies impact your business model. *Business Horizons*, *62*(3). [Online]. Available: https://doi.org/10.1016/j.bushor.2019.01.009

[2] Zheng, Z., Xie, S., Dai, H., Chen, X.; Wang, H. (2017). An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends. Proceedings - 2017 IEEE 6th International Congress on Big Data, BigData Congress 2017, 557–564.

[3] Chang, V., Baudier, P., Zhang, H., Xu, Q., Zhang, J.; Arami, M. (2020). How Blockchain can impact financial services – The overview, challenges and recommendations from expert interviewees. Technological Forecasting and Social Change, 158. [Online]. Available: https://doi.org/10.1016/j.techfore.2020.120166

[4] Cao, Y., Li, Y., Coleman, S., Belatreche, A., Mcginnity, T. M. (2016). Detecting Wash Trade in the Financial Market. [Online]. Available: https://ieeexplore.ieee.org/document/6924058

[5] Akartuna, E. A., Johnson, S. D., Thornton, A. E.: (2022). The money laundering and terrorist financing risks of new and disruptive technologies: a futures-oriented scoping review. Security Journal. [Online]. Available: https://doi.org/10.1057/s41284-022-00356-z

[6] Ali, A., Abd Razak, S., Othman, S. H., Eisa, T. A. E., Al-Dhaqm, A., Nasser, M., Elhassan, T., Elshafie, H., & Saif, A. (2022). Financial Fraud Detection Based on Machine Learning: A Systematic Literature Review. In Applied Sciences (Switzerland) (Vol. 12, Issue 19). MDPI. [Online] Available: https://doi.org/10.3390/app12199637

[7] Baesens, B., Van Vlasselaer, V., & Verbeke, W. (2015). Fraud analytics using descriptive, predictive, and social network techniques: a guide to data science for fraud detection. John Wiley & Sons.

[8] Zheng, W., Zheng, Z., Dai, H. N., Chen, X., Zheng, P.: (2021). XBlock-EOS: Extracting and exploring blockchain data from EOSIO. Information Processing and Management, 58 (3). [Online]. Available: https://doi.org/10.1016/j.ipm.2020.102477

[9] Bank, D., Koenigstein, N., Giryes, R.: (2020). Autoencoders. [Online]. Available: http://arxiv.org/abs/2003.05991

[10] Hariri, S., Kind, M. C., & Brunner, R. J. (2021). Extended Isolation Forest. IEEE Transactions on Knowledge and Data Engineering, 33(4). [Online]. Available: https://doi.org/10.1109/TKDE.2019.2947676M

[11] He, Z., Xu, X., Deng, S.: (2003). Discovering cluster-based local outliers. Pattern Recognition Letters, 24 (9–10), 1641–1650. [Online]. Available: https://doi.org/10.1016/S0167-8655(03)00003-5

Federated Learning for Edge Computing for Industry 4.0

[1]*Alexander BRECKO (2nd year),*
Supervisor: [2]Iveta ZOLOTOVÁ Consultant: [3]Erik KAJÁTI

[1,2,3]Dept. of Cybernetics and Artificial Intelligence, FEEI TU of Košice, Slovak Republic

[1]alexander.brecko@tuke.sk, [2]iveta.zolotova@tuke.sk, [3]erik.kajati@tuke.sk

***Abstract*—With the rise of the Internet of Things paradigm, the number of devices and the data they generate is growing exponentially. To handle this growth, computing moves from the cloud to the network's edge. However, computing resources are, in many cases, limited at the network's edge. In this paper, we provide a brief overview of techniques that could enable data diversification across multiple edge devices that would participate in training a global model from various data sources using federated learning. The paper also summarizes work done over the last year.**

***Keywords*—Edge AI, edge computing, federated learning, Internet of Things**

I. Introduction

In recent years, the Internet of Things (IoT) has expanded rapidly, connecting many devices to the Internet and enabling sensing and computing. Much of the computing has been done using cloud services, but recently there has been an increasing emphasis on edge computing. Given the power of these edge devices, it is possible to move computation and processing from the cloud to the edge, as well as simple artificial intelligence (AI) and machine learning (ML) tasks [1]. While there is no denying that ML is on the rise, there may be severe privacy concerns. The biggest problem is that millions of images, videos, or data are collected centrally on a single server. In our research, we are primarily concerned with training models at the edge of the network. Closely related to this is federated learning (FL), where clients take care of the computation rather than one central server. FL is an approach that gained awareness around 2016 and 2017 when Google introduced the concept of the Google keyboard (Gboard) [2].

II. The initial status

At last year's conference, we provided an overview of edge devices that could gather data from sensors and actuators. An analysis of edge devices was also performed in paper [3], where we focused more explicitly on devices that can deploy AI at the edge of the network and on training AI at the network's edge. In this case, we focus primarily on devices with an ARM architecture that are commercially available and have a maximum power consumption of up to 40W. In this case, several devices were examined, such as NVIDIA Jetson Family, Huawei Atlas Family, Coral Family, Bearkey, Bitmain, Ultra96-V2, and other boards. Among the devices examined are Intel Neural Compute Stick Family, Google Coral USB Accelerator, Orange Pi AI Stick Lite, or Rockchip RK1808 AI Compute Stick. We would like to focus on interfacing the edge devices aforementioned with AI using FL.

FL is a technique for developing ML models utilizing data distributed across several devices, such as mobile phones and smart home devices [4]. The concept of FL is to train an ML model using the data saved locally on each device instead of centralizing all data in one location. In this case, the devices train their local models, whose weights are then sent to a server where they are aggregated. This way, a global model is created [5].

III. The tasks solved in the previous year

In our research, we have focused on FL techniques that can be used with edge computing. As part of our research, a review paper [6] has been published, focusing exactly on FL techniques with computations at the network's edge.

As a first step, we each focused on deploying AI at the network's edge. In this case, there are several options. We decided to test the different edge devices available to our research group. As we would like to focus on classification and detection in the future, in this case, we have chosen a pre-trained model. Therefore, we have selected the YOLOv5 (You Only Look Once version 5) model for this case [7]. YOLOv5 is available in several versions, including nano, small, medium, large, and extra large. Each of these versions needs different hardware. As part of deploying AI at the network's edge, we conducted many tests to determine if the model could be executed on the device. Specifically, it concerns detecting and classifying vehicles from an IP camera at a highway overpass. This camera provides a stream with 25 frames per second (FPS), to which we can connect and process individual frames. When we were able to run the model, we observed how smoothly the program ran and how many FPS the device could process. These experiments are briefly summarized in Table I. These experiments were performed on 4 different devices and show that today's edge devices can run applications using a simple model that are not very computationally intensive. Specifically, these are an NVIDIA Jetson Nano, a Raspberry Pi 4, a Macbook Air M2, and a server computer with an NVIDIA RTX Titan graphics card. As you can see from the table, Jetson Nano and RPI4 could only run the nano and small models. The larger model could not be uploaded to the device because it needed more than 4GB of memory. Although the Macbook Air M2 is also an ARM architecture device, this device could also run the medium size model because the memory size, in

this case, is 16GB. We also compared these experiments with a high-performance server computer with an NVIDIA RTX Titan graphics card with 24GB of memory. In this case, it was possible to run even the largest model.

TABLE I
YOLOv5 EXPERIMENTS

YOLO	Jetson Nano	RPI4	Macbook M2	RTX Titan
Nano	11-13 FPS	8-10 FPS	23-25 FPS	100+ FPS
Small	6-9 FPS	3-5 FPS	12-15 FPS	100+ FPS
Medium	-	-	6-7 FPS	80-85 FPS
Extra large	-	-	-	33-36 FPS

As we need to diversify computation and offload clouds, we are starting to discuss training models on edge devices. Security and privacy are closely related to this. In classical training, it is necessary to have data collected on a central server. In this case, we can deal with training models using FL. In this solution, we tried to run FL training on edge devices. Within our already mentioned article [6] was also made an overview of frameworks that can support FL on ARM architecture. The available frameworks for FL were Flower [8] and TensorFlow Federated [9], but since TensorFlow Federated is currently only in beta, we selected Flower, which provides excellent documentation. Using this framework, it can perform FL training on multiple devices simultaneously. In this case, we tried different training combinations to involve as many devices as possible. Our primary objective will be the classification of vehicles, where we need to classify objects into 10 classes (motorcycle, car, car with trailer, van, van with trailer, small truck, large truck, large truck with trailer, bus, unknown object).

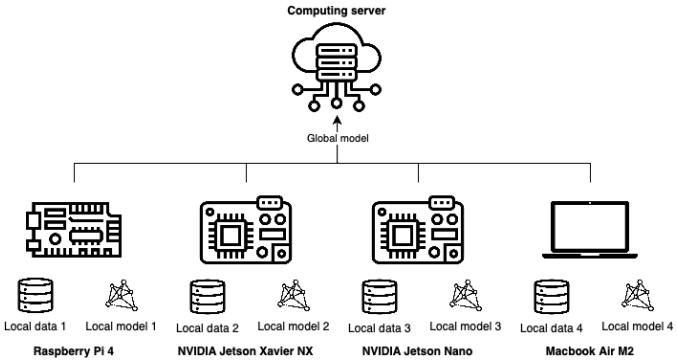


Fig. 1. Architecture of FL

For the server, it is recommended to use a more powerful device with a good internet connection because it takes care of all the communication of the individual devices and the aggregation of the global model. As a server, we chose the already mentioned PC, which contains NVIDIA RTX Titan GPU. Clients can be diverse in these cases. These can be heterogeneous devices with different hardware and software in a real deployment. In the current experiments, we are focusing primarily on ARM architecture. These are NVIDIA Jetson Nano and Jetson Xavier NX, Raspberry Pi, and Macbook Air M2. The simple architecture of our problem is shown in Figure 1, where the individual edge devices are shown with their private data. These experiments are still in an early phase, but in the first experiments, we achieved an accuracy of about 85%. This accuracy can undoubtedly be improved. We need to find the optimal setting of the whole communication, the number of rounds that should take place between the clients and the server, and also the number of epochs that run on individual devices. When we train, we use TensorFlow libraries, which each have different versions on the devices. These details still need to be fine-tuned because they cause problems in calculations and training. In the case of the Jetson Nano, this causes problems, and it is not yet possible to use the full potential of the hardware it has.

IV. Future work

Further, we plan to develop a methodology for using FL to support Industry 4.0 and transport using edge AI devices. We would like to focus on allowing each device to use its maximum potential performance (calculations using the GPU, not just the CPU) and also to optimize the whole communication and training of the model (number of rounds to be executed in FL, number of epochs, complexity of local models) for the deployment of the model in practice. In this case, we would like to leverage and implement strategies and methods, such as FedProx, that can adapt to different devices with different computational capabilities and different data heterogeneity. This will allow each device to utilize its maximum potential performance by utilizing both the GPU and CPU for efficient computations.

Acknowledgment

This publication was supported by the APVV grant ENISaC - Edge-eNabled Intelligent Sensing and Computing (APVV-20-0247)

References

[1] T. Mazhar, H. M. Irfan, I. Haq, I. Ullah, M. Ashraf, T. A. Shloul, Y. Y. Ghadi, Imran, and D. H. Elkamchouchi, "Analysis of challenges and solutions of iot in smart grids using ai and machine learning techniques: A review," *Electronics*, vol. 12, no. 1, 2023. [Online]. Available: https://www.mdpi.com/2079-9292/12/1/242

[2] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y. Arcas, "Communication-efficient learning of deep networks from decentralized data," 2016. [Online]. Available: https://arxiv.org/abs/1602.05629

[3] L. Pomšár, A. Brecko, and I. Zolotová, "Brief overview of edge ai accelerators for energy-constrained edge," in *2022 IEEE 20th Jubilee World Symposium on Applied Machine Intelligence and Informatics (SAMI)*. IEEE, 2022, pp. 000 461–000 466.

[4] L. Li, Y. Fan, M. Tse, and K.-Y. Lin, "A review of applications in federated learning," *Computers & Industrial Engineering*, vol. 149, p. 106854, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360835220305532

[5] H. R. Roth, K. Chang, P. Singh, N. Neumark, W. Li, V. Gupta, S. Gupta, L. Qu, A. Ihsani, B. C. Bizzo, Y. Wen, V. Buch, M. Shah, F. Kitamura, M. Mendonça, V. Lavor, A. Harouni, C. Compas, J. Tetreault, P. Dogra, Y. Cheng, S. Erdal, R. White, B. Hashemian, T. Schultz, M. Zhang, A. McCarthy, B. M. Yun, E. Sharaf, K. V. Hoebel, J. B. Patel, B. Chen, S. Ko, E. Leibovitz, E. D. Pisano, L. Coombs, D. Xu, K. J. Dreyer, I. Dayan, R. C. Naidu, M. Flores, D. Rubin, and J. Kalpathy-Cramer, "Federated learning for breast density classification: A real-world implementation," in *Domain Adaptation and Representation Transfer, and Distributed and Collaborative Learning*, S. Albarqouni, S. Bakas, K. Kamnitsas, M. J. Cardoso, B. Landman, W. Li, F. Milletari, N. Rieke, H. Roth, D. Xu, and Z. Xu, Eds. Cham: Springer International Publishing, 2020, pp. 181–191.

[6] A. Brecko, E. Kajati, J. Koziorek, and I. Zolotova, "Federated learning for edge computing: A survey," *Applied Sciences*, vol. 12, no. 18, 2022. [Online]. Available: https://www.mdpi.com/2076-3417/12/18/9124

[7] Yolov5 documentation. [Online]. Available: https://docs.ultralytics.com/

[8] Flower a friendly federated learning framework. [Online]. Available: https://flower.dev

[9] Tensorflow federated: Machine learning on decentralized data. [Online]. Available: https://www.tensorflow.org/federated

Characterization and structural studies of magnetic fluids

[1]*Maksym KARPETS (4th year)*
Supervisor: [2]Milan TIMKO

[1]Dept. of Physics, FEI TU of Košice, Slovak Republic

[1,2]Institute of Experimental Physics SAS, Košice, Slovak Republic

[1]maksym.karpets@tuke.sk, [2]timko@saske.sk

***Abstract*— This paper is a summarisational article in the range of 2 pages, which is a brief overview of my work for the past year of study. The main research task is aimed on characterization of ferrofluids (FFs) and experimental investigation of the magnetic fluid-solid interfaces under the influence of electric field.**

***Keywords*—external fields, magnetic fluids, magnetic nanoparticles, transformer oil-based ferrofluids.**

I. INTRODUCTION

Magnetic fluids (or ferrofluids) are suspensions of magnetic nanoparticles (MNPs) in a liquid carrier [1]. The understanding and development of the mechanisms of NP assembly and manipulation of the assembly process are highly important in the synthesis of the well-defined architectures at a nanometer scale [2]. In general, the stimulation and fabrication of novel functional surface nanostructures opens horizon for new materials and devices. The NP assembling could lead to controllable arrangement in bulk or at interfaces and it can be achieved just by using some template or external stimuli such as magnetic or electric fields. In this regard, self-assembly which provides one of the cheapest and easiest ways for organizing NP arrangements into well-ordered structures is of current interest. Particularly MNPs have gained great interest in last decade owing to their applications in medicine, cancer theranostics, biosensing, catalysis, agriculture, electrical engineering and the environment [3]. Especially assembling of MNPs on planar interfaces from liquid dispersions took much attention.

II. INITIAL STATUS

The electric field-controllable properties of dielectric FFs were recently reported [4]. For such ferrofluids, the adsorption can be enhanced by applying non-homogeneous magnetic fields perpendicular to the interface [5]. Based on the small-angle neutron scattering (SANS) studies of a transformer oil-based ferrofluid (TOFF), the bulk structuring and phase separation under electric fields was revealed. The nanofluids are of practical importance - especially they have been used in several thermal management systems, as they contribute to the augmentation of the heat dissipation rates in many applications. For instance, in high-voltage engineering TOFF serve as an insulating and cooling additive in power transformers [6]. It was concluded that the electric field causes electrical polarization of the particles, so that electric dipole-dipole interaction dominates and acts in a similar way like external magnetic field. The theoretical consideration [7] reveals that sufficiently strong applied electric fields cause non-uniform distributions of particles between electrodes. The macroscopic structural changes in a simple ferrofluid consisting of iron oxide nanoparticles coated with oleic acid and dispersed in transformer oil under electric field have been proven [8]. Also, an increase in the ferrofluid viscosity with increasing electric field intensity (that is analogous to the magnetoviscous effect) is associated with the formation of elongated chain-like aggregates. The electric response of ionic ferrofluids to an ac electric field was experimentally investigated by means of the impedance spectroscopy technique [9].

III. WHAT I HAVE SOLVED THIS YEAR

A. Characterization of magnetic fluids

The investigated MFs are based on a commercially available inhibited insulating transformer oils - SHELL Diala S4 ZX-I and TO MOL 40A. Magnetite and $Mn_{0.6}Zn_{0.4}Fe_2O_4$ MNPs were synthesized by chemical co-precipitation method from aqueous solution of ferrous and ferric ions in the presence of NH_4OH at 80–82 °C. After coprecipitation, the MNPs were sterically stabilized by chemisorbing of a single oleic acid layer ($C_{18}H_{34}O_2$, Merk) on the particle surfaces. Synthesis of FF was performed according to previous reports [10]. Samples of ferrofluids with a low and a high NP concentration were prepared.

The ferrofluids were characterized by magnetization curves, magnetic susceptibility and temperature-dependent magnetization measurements [11]. The nanoparticle size distribution is determined from dynamic light scattering and the magnetization data. The magnetic properties of the MF sample were measured by means of a vibrating sample magnetometer installed on a cryogen-free superconducting magnet from Cryogenic Limited (IEP SAS, Kosice). The obtained magnetization curves were measured at 5 K and 300 K in the field ranging up to 6 T (saturation is at 2 T). The mean particle size derived from the fitting of the magnetization curve by the superposition of Langevin functions is 11.4 nm. On the other hand, the mean hydrodynamic nanoparticle diameter is 19.5 nm, as revealed by the dynamic light scattering (DLS) method (Malvern Instruments Ltd.,Malvern, UK). AC magnetic susceptibility was measured by an AC susceptometer (IMEGO, DynoMag,

SE). Analogously to the AC magnetic susceptibility spectra, the real part of dielectric permittivity and dissipation factor of the oil and the ferrofluid samples in dependence on the electric field frequency were measured. The measurements were performed by means of an LCR meter (Agilent E4980A).

B. Reflectometry study of magnetic fluids under electric field

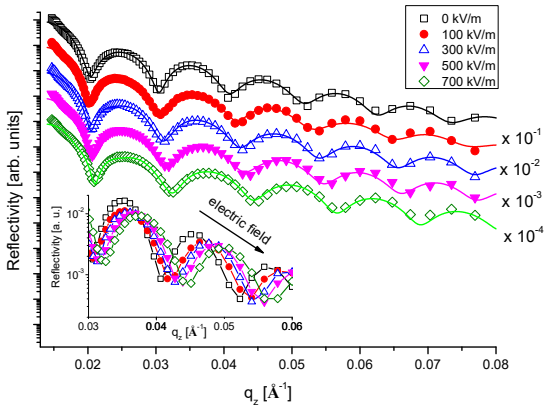


Fig. 1. Specular reflectivity curves obtained in the NR experiment when applying electric field from 0 to 700 kV/m. Solid lines represent best fits [2].

Specular reflectivity of non-polarized neutrons was measured at the neutron reflectometer GRAINS with a horizontal sample plane configuration installed at the IBR-2 pulsed reactor of JINR (Dubna) [2]. The measurements were carried out in the 'beam from the bottom' configuration in the time-of-flight mode that is standard positioning to study the solid-liquid interfaces when a plane neutron beam passes through a single-crystal silicon substrate (thickness of about 15 mm) to meet and scatter from the interface between a liquid sample and an electrode film (thickness below 1000 Å) deposited prior the experiment on the substrate. The sample in the experimental cell was exposed to direct current (DC) electric field of various intensities (from 100, 300, 500 and 700 kV/m), which are close to the operating values commonly used in high voltage power engineering. The obtained experimental and best model NR curves for different electric field intensity applied are presented in Fig. 1. The reflectivity curves for the interface under electric field were fitted with all fixed values of the solid components parameters and all varying parameters of FF components. The received fit parameters from the previous dataset (with smaller electric field) are used as a starting point for the next dataset (with higher electric field). The electric field induced evolution of the 'liquid part' of the interface in terms of the changes in the SLD profile is followed in Fig. 2. As expected, the increase in the electric field slightly affects the structure of layer #1 determined by the surfactant shell; one can see small reduction in thickness, roughness and SLD of this layer. More significant changes are observed for layer #2. It becomes thicker and more concentrated thus evidencing the enhanced adsorption of MNPs. The layer thickness growths with the electric field increase together with the SLD. Starting from 300 kV/m, the fits of the reflectivity curves are better if an additional layer #3 is introduced. It describes a transition to the bulk ferrofluid. At the same time, some redistribution of MNP in the two layers takes place at maximum field intensity, reflecting a decrease of the MNPs content in the first layer. The reason of such development is the polarization of the particles and their interaction as dipoles. Moreover, from the relaxation measurements we can conclude that a residual electric field and potentially present magnetic and Van der Waals forces influence nanoparticles assembling even some time after switching the field off. The observed self-assembled layering could be used as an additional barrier at the inner surface of transformer to increase dielectric breakdown voltage of working fluids.

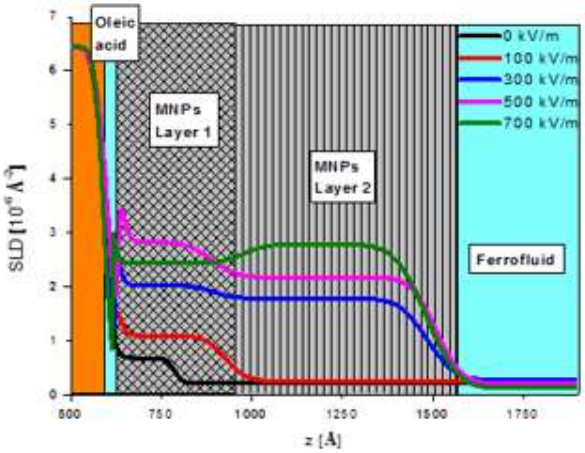


Fig. 2. SLD profiles at the electrode-ferrofluid interface at different intensities of the external electric field [2].

IV. Conclusion and Next Steps

We report assembling of superparamagnetic nanoparticles of a dilute classical ferrofluid - MNPs coated with oleic acid in transformer oil. Forming of magnetic nanoparticles layers was found on a planar surface of copper electrode film on a single-crystal silicon substrate under perpendicular homogeneous electric field by neutron reflectometry. As the next step for using of such materials in electric power engineering investigation of magnetic field and temperature effect on FF is needed. We also have accepted proposals for neutron measurements for other FFs - nanomagnetite into organic solvents and TOFF with adding fullerenes.

Acknowledgment

The work is a part of the projects supported by the Programme of SAS grants for PhD students (No. APP0140); the Science Grant Agency of the Slovak Republic (No. 2/0011/20 and 2/0016/17); the Slovak Research and Development Agency (No. APVV-18-0160).

References

[1] S. Odenbach, Colloidal magnetic fluids: basics, development and application of ferrofluids, Vol. 763, Springer Berlin Heidelberg, (2009).
[2] M. Karpets et al., Journal of Molecular Liquids 362, 119773 (2022).
[3] K. Theis-Bröhl, et al., Nanomaterials. 10, 1231 (2020).
[4] M. Rajnak et al., Appl. Phys. Lett. 107, 073108 (2015).
[5] A. Nagornyi et al., "Particle assembling induced by non-homogeneous magnetic field at transformer oil-based ferrofluid/silicon crystal interface by neutron reflectometry", Appl. Surf. Sci. 473, 912, 2019.
[6] M. Rajnak et al., Energies 12, 4532, (2019).
[7] P.A. Selyshchev, et al., J. Mol. Liq. 278 491–495, 2019.
[8] M. Rajnak et al., "Structure and viscosity of a transformer oil-based ferrofluid under an external electric field", J. Magn. Magn. Mater. 431 99–102, 2017.
[9] F. Batalioto et al., Journal of Electroanalytical Chemistry 874, 114452 (2020).
[10] M. Karpets et al., Ukr. J. Phys. 65, 729 (2020).
[11] M. Rajňák et al., J. Phys. D: Appl. Phys. 55, 345002 (2022).

Biomedical and biotechnological applications of magnetic nanoparticles

[1]*Kristina ZOLOCHEVSKA (2nd year)*
Supervisor: [2]Peter KOPČANSKÝ

[1,2]Institute of Experimental Physics of the Slovak Academy of Sciences, Košice, Slovak Republic

[1]kristina.zolochevska@tuke.sk, [2]kopcan@saske.sk

***Abstract*— Many pathological phenomena in the body are associated with structural changes in the storage protein of iron - ferritin. Magnetite formation has been observed in people with neurodegenerative, cardiovascular, and oncological diseases. The cause has not yet been fully elucidated. Particularly incurable neurodegenerative diseases are also associated with the formation of amyloid plaques, similar to fibrillar protein structures composed primarily of beta-pleated leaves. Ferritin's cavity can be used to make a material known as magnetoferritin. The iron-based magnetic nanoparticles present in magnetoferritin give a stronger sensitivity to the applied magnetic field, making it possible to investigate the hyperthermic effect. With the help of magnetic nanoparticles and ferritin derivatives, it became possible to diagnose and treat tumors, as well as deliver drugs to the right parts of the body due to its peroxidase-like catalytic activity. The main goal of this proposed work is to do the experimental study of impact of external fields on magnetic nanoparticles and to study the influence of electromagnetic radiation on ferritin and its derivatives.**

***Keywords*— Biomedical applications, magnetic nanoparticles, ferritin.**

I. Introduction

Iron is the one of essential elements for living organisms. It generally exists in two different oxidation states. The first is a relatively soluble but highly toxic ferrous form (Fe^{2+}), and the second is a very insoluble but nontoxic ferric state (Fe^{3+}). Both iron deficiency and excessive iron levels can be harmful, and thus, it is stored in the interior of a spherical protein called ferritin, which is becoming increasingly significant in biomedicine.

Ferritin is a spherical cage that is composed of an apoferritin shell and iron-based magnetic nanoparticles, which surrounds ferrihydrite (FeOOH) nanocrystals. The core of native ferritin might turn from ferrihydrite to magnetite, thereby forming biogenic magnetoferritin [1]. Magnetoferritin is comprised of an apoferritin shell and iron-based magnetic nanoparticles, which results in increased sensitivity to an applied magnetic field [2].

Magnetic nanoparticles are well-regarded for their low toxicity, biocompatibility, and high surface area, as well as their intrinsic magnetic properties. An effective surface engineering of magnetic nanoparticles can be used as a strategy to precisely adjust its ability to target specific cells and enhance its intrinsic properties and other crucial features for biomedical applications such as colloidal stability [3]. In this context, numerous surface engineering methods for magnetic nanoparticles have been employed to improve modern therapeutic approaches, such as in drug delivery systems, magnetic resonance imaging (MRI), and hyperthermia.

The objectives of the work are:

1. To make synthesis of magnetoferritin with different loading effects (to investigate the dependencies of concentration of iron atoms in protein biomacromolecules on x-ray patterns).
2. To study the interaction between magnetic nanoparticles (including magnetoferritin) and lysozyme amyloid fibrils in a magnetic and rotating field and reveal the possible mechanism of pathology initiation or the potential therapeutic effect.

II. Analysis of the topic

Magnetic fluids

Magnetic fluids (MF), also known as ferrofluids, are highly stable colloidal suspensions, which are made up of magnetic nanoparticles, coated with a surfactant and dispersed in a carrier fluid, either water or oil. The most studied and advanced ferrofluids are based on iron oxide nanostructures especially nanoparticles, because of their easy and large-scale synthesis at low costs [4]. The focus in this area is mainly on the possibility of interaction between a magnetic fluid and a magnetic field. The magnetic properties of magnetic fluids allow for the precise and controlled delivery of drugs to specific targets in the body. By exposing the ferrofluid to a magnetic field, the movement of the nanoparticles can be controlled, enabling the drug to be directed to the desired location.

Magnetic nanoparticles

Magnetic nanoparticles are responsible for the magnetic properties in magnetic fluids. Therefore, the next chapter is devoted to the structure and magnetic properties of magnetic nanoparticles. Magnetic nanoparticles (MNPs) are nanoscale particles (1–100 nm) that can be guided through an external magnetic field due to their superparamagnetic, ferrimagnetic, and ferromagnetic properties. There are known many phases of iron oxide MNPs, but magnetite (Fe_3O_4) and maghemite (γ-Fe_2O_3) are the most commonly used because of their easy surface modification by organic or inorganic compounds.

Magnetic properties of nanoparticles. For applications that require high performance, the ideal nanoparticles should have a spherical structure, be biocompatible, possess superparamagnetic properties, have a narrow size distribution, high crystallinity, a large surface area and good dispersion in

liquid medium. The relaxation time of magnetization in response to external magnetic fields in magnetic particle imaging (MPI) and magnetic particle spectroscopy (MPS) is established by the Brownian and Néel relaxation processes [5].

Magnetoferritin

Magnetoferritin is a magnetic nanomaterial with good biocompatibility and flexibility for studying magnetite bio-mineralization within a controlled and uniform cavity. The magnetic nanoparticles are prepared by adding Fe/oxidant to the apoferritin cage under conditions of 65°C, alkaline pH, constant stirring. The amount of iron atoms added to the reaction impacts the size of the iron core, the occurrence of aggregates, and its chemical composition. Magnetoferritin forms through a process involving the entry of Fe^{2+} ions into the apoferritin shell via hydrophilic channels under the influence of an electrostatic gradient. The Fe^{2+} ions are then oxidized to Fe^{3+}, followed by nucleation and growth of the inorganic core.

III. Summary

Preparation. At the moment, the synthesis of magnetoferritin with different loading effects has been carried out. This experiment was done following the instructions above, but the loading effects were changed.

Fig. 1. The final product of the synthesis: magnetoferritin colloidal solution with protein concentration with loading ~ 550 Fe atoms per the one protein biomacromolecule; and higher iron loading ~ 730 Fe atoms per the one protein biomacromolecule; and the highest iron loading ~ 810 Fe atoms per the one protein.

As a result, a sediment was detected at a high loading effect. Sediment depends on loading factor. With higher iron loading, sedimentation is higher. About 500 and more iron atoms pre one protein led to higher sizes, and after time we can observe sediment. Protein shell (apoferritin) destroying depends on the higher iron loading, higher temperature, or extremely pH changes. Addition of NaOH can usually partially stabilize sample and led preferable to magnetite formation. Turbidity can be in this way eliminated.

Using. Besides iron accumulation and mineralization of iron oxides nanoparticles, neurodegenerative disorders are characterized by the presence of abnormal protein aggregates (amyloid plaques). Our main goal was to study the interaction between ferritin derivatives and lysozyme amyloid fibrils in the present work. To accomplish that, we have successfully synthesized magnetoferritin (MF) and reconstructed ferritin (RF) (with low loading factors) as well as lysozyme amyloid fibrils (from native lysozyme) using controlled in vitro synthesis. To reveal the potential adverse effect of ferritin derivatives with lysozyme amyloid fibrils interaction, we determined the time dependence of ferrous ions released from the ferritin envelopes.

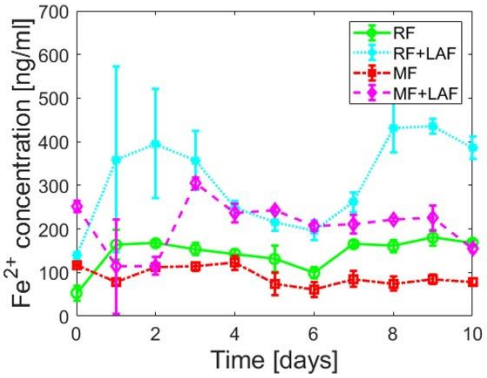


Fig. 2. Time dependence of ferrous ion release (mean ± Standard deviation) from ferritin's derivatives and during the interaction with Lysozyme amyloid fibrils.

The analysis of the spectrophotometric results (Figure 2) showed an increase in the reduction of the mineral core after interaction with LAF. An extended monitored incubation time of 11 days was chosen to examine iron release kinetics in vitro. The results suggest that ferrous ion scavenging is a spontaneous and energetically favorable process at room temperature. Wave trends probably matched iron reoxidation and reuptake by ferritin derivatives, i.e., iron flow. The current results support the idea that the main factor contributing to fibril elimination is the redox potential of the iron core, i.e., electron transfer during the reduction of ferritin derivatives from Fe^{3+} to Fe^{2+}.

Ferrous ions are highly toxic to the cell. It is clear that the release of toxic ferrous ions occurs to a greater extent during the interaction of ferritin's derivatives with lysozyme amyloid fibrils. Induction of increased ferrous ions release from ferritin's envelope during interaction with lysozyme amyloid fibrils (LAF) can cause increased oxidative stress and more significant damage to the cell.

Another goal of the work will be to study the interaction between magnetic nanoparticles and lysozyme amyloid fibrils in a magnetic and rotating field, to measure the influence of magnetic nanoparticles of various forms in the field of magnetic hyperthermia, and to study the interaction between magnetic nanoparticles and polychlorinated biphenyl (PCB) to examine the reduction of PCB toxicity.

References

[1] L. Xue, D. Deng and J. Sun, "Magnetoferritin: Process, Prospects, and Their Biomedical Applications," *International Journal of Molecular Sciences,* vol. 20, no. 10, p. 2426, 2019.

[2] L. Balejcikova, M. Molcan, J. Kovac, M. Kubovcikova, K. Saksl, Z. Mitroova, M. Timko and P. Kopcansky, "Hyperthermic effect in magnetoferritin aqueous colloidal solution," *Journal of Molecular Liquids,* vol. 283, pp. 39-44, 2019.

[3] S. M. Silva, R. Tavallaie, L. Sandiford, R. D. Tilley and J. J. Gooding, "Gold coated magnetic nanoparticles: from preparation to surface modification for analytical and biomedical applications," *Chemical Communications,* vol. 52, pp. 7528-7540, 2016.

[4] M. Imran, A. M. Affandi, M. M. Alam, A. Khan and A. I. Khan, "Advanced biomedical applications of iron oxide nanostructures based ferrofluids," *Nanotechnology,* vol. 32, no. 42, 2021.

[5] Jeun, Minhong, K. Y. Jeong, P. K. Ho, P. S. Ha and B. Seongtae, "Physical Contribution of Néel and Brown Relaxation to Interpreting Intracellular Hyperthermia Characteristics Using Superparamagnetic Nanofluids," *Journal of Nanoscience and Nanotechnology,* vol. 13, no. 8, pp. 5719-5725, 2013.

Load Distribution in Container-based Edge Computing

[1]*L'ubomír URBLÍK (1st year),*
Supervisor: [2]*Peter PAPCUN*

[1,2]Dept. of Cybernetics and Artificial Intelligence, FEEI TU of Košice, Slovak Republic

[1]lubomir.urblik@tuke.sk, [2]peter.papcun@tuke.sk

***Abstract*—This article discusses the various methods used for load distribution in multi-node solutions. The rise of the edge computing paradigm in Internet of Things solutions brings forward new challenges, requiring improvements to the existing methods. As the devices are often constrained by their size or processing power, it is important to distribute the load efficiently to take full advantage of the edge computing. Due to the heterogeneity of the devices used, a unifying platform like Docker provides an advantage when creating such solutions. Artificial intelligence can be advantageous as it can provide intelligence for load distribution methods and increase performance, adaptability, and resilience. The next step will be testing different artificial intelligence methods in combination with various load distribution methods to take full advantage of the processing power available at the edge.**

***Keywords*—containerization, edge computing, intelligent load distribution, internet of things**

I. Introduction

In recent years, the Internet of Things (IoT) saw a massive increase in the number of devices connected to the Internet. The onset of autonomous vehicles, which generate and process extensive amounts of data, brought new requirements for the IoT infrastructure. The need to process this data quickly and efficiently, which allows the vehicle to make split-second decisions, is something not achievable by only using cloud computing. Places like tunnels or roads in the middle of nowhere do not offer great connectivity, which poses a problem for communication with the cloud. The edge computing paradigm solves this problem, as the data processing takes place near the data source, where it is impervious to such problems. As the processing power of edge devices cannot match that of cloud data centers, a new question arises - is it better to process the data locally, on a slower machine, or in the cloud, where the distance adds latency? To answer this question, we have looked at multiple solutions allowing the processing to be distributed between the nodes located at the network's edge and between the edge and the cloud, depending on different conditions and factors.

II. IoT infrastructure

Traditional IoT solutions follow the same three-layered infrastructure - edge, fog, and cloud - with things at the zeroth level [1]. This architecture is shown in Fig. 1.

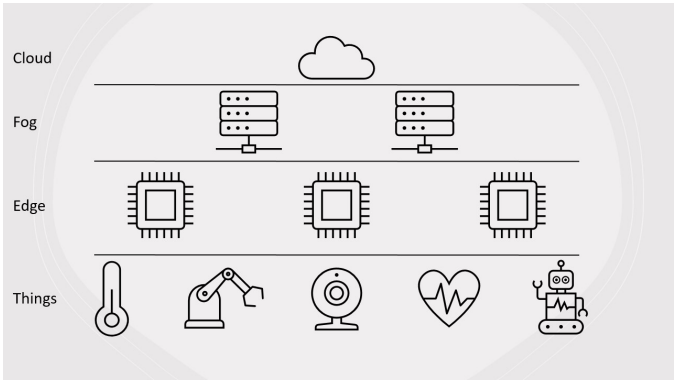


Fig. 1. Traditional three-layered IoT architecture, made up of edge, fog, and cloud

A. The things

The lowest tier of the infrastructure represents the physical devices. These devices collect data from their surrounding environment or are used to influence the physical world around them. This term encompasses a wide variety of devices, for example, wearables, such as smartwatches, which monitor the wearer's heartbeat, temperature sensors used to control a building's heating and cooling system, or vibration sensors used in manufacturing for preventative maintenance. Such devices contain a microcontroller unit (MCU), single-board computer (SBC), sensors, and/or actuators [2].

B. The edge

The edge represents the boundary between the local network and the Internet. This can mean either the first device, which can connect to the Internet, or a more powerful device in proximity. For this publication, we will use this term to represent devices available in the local network, such as local servers or SBCs to which MCUs are connected. There are numerous advantages to processing the data at the edge, the first of which is privacy. The resiliency and security increase as the data does not leave the local network. This protects the data from being accessed by unwanted actors, such as hackers or blackmailers. Another advantage is the reduced latency, as the data has a much shorter distance to travel compared to being sent to the cloud. When used in conjunction with cloud computing, pre-processing the data at the edge will decrease costs, and the amount of resources and storage required [1].

C. The fog

Placed between the edge and the cloud, the fog allows us to further process our data before sending it to the cloud. By utilizing traditional servers, the fog substantially increases processing power compared to the edge. This layer also allows for data storage, such as storing daily reports before sending monthly reports to the cloud [1].

D. The cloud

The cloud represents resources available wherever and whenever through the Internet. Unlike edge, which is highly decentralized, cloud computing resources are centralized in large data centers. These centers contain hundreds of high-power physical servers, further divided into virtual machines. One of the most efficient methods is containerization, in which apps are divided into separate containers, isolated from each other. Traditional virtualization methods require a client operating system, which can lead to compatibility issues with both hardware components and software. Docker, a containerization platform, does away with this requirement and allows for a "write once, run anywhere" approach. This ensures that any application created for Docker will run on any device capable of running Docker. The cloud offers unmatched computing performance, but it also has its drawbacks. The distance from the devices alone brings a high increase in latency. Best suited for long-term storage or time-insensitive data processing, the cloud represents the final destination of most data collected by the devices [3].

III. Current state of edge computing

Edge computing presents many challenges for device manufacturers, as the users often require the devices to be small and consume little power while still offering good connectivity and processing power. These devices can be divided into three categories,

Microcontroller units, the weakest edge devices, represent the simplest and cheapest available options. These devices are most often used to connect the sensors and actuators to collect or generate the data. The data collected is then sent upstream to a more powerful device, where it is processed. Typical examples include platforms like Arduino, which are popular for do-it-yourself solutions due to their low-cost and accessibility to new programmers [4].

Single board computers, the middle of the pack, contain stronger hardware and offer more flexibility in data processing or programming languages when developing a new solution. Devices like Raspberry Pi or NVIDIA Jetson offer an inexpensive way to run more complex data processing at the edge. Unlike MCUs, SBCs can run fully featured operating systems and are therefore often capable of running a wider variety of programs [5].

Most often incorporating the traditional x86 CPU architecture, as well as ARM in recent years, the mini PCs are more akin to traditional PCs than to specialized edge devices. Their size allows them to be deployed to a space-constrained environment, such as rooms or vehicles. They also offer expandability, allowing for hardware upgrades if needed. This makes them a flexible choice at a somewhat higher price [6].

IV. Traditional load balancing methods

When using multiple processing nodes, there are various methods to increase their performance, the stability of the system and decrease the latency or energy consumption. The first of such methods is load balancing, the technique of swapping which node processes which request. This process tries for the processing to be split somewhat evenly, ensuring no node is overutilized or underutilized. This would lead to an increase in processing time and energy consumption and might lead to data loss.

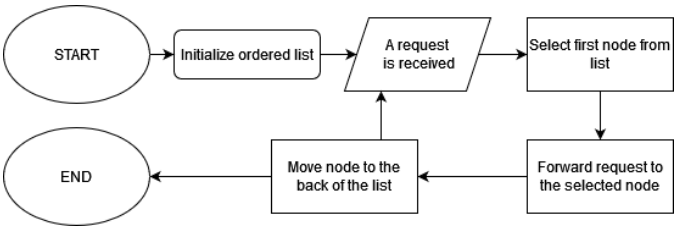


Fig. 2. Flowchart of the round-robin algorithm

In the *round-robin algorithm*, requests are sent to nodes in a predefined order. The process starts with an ordered list of nodes. When a request arrives, it is sent to the first node in the list. This node is then moved to the tail of the list. The flowchart of this algorithm is shown in Fig. 2. The main advantage is the ease of implementation and no resources sitting idle. The disadvantages include uneven distribution, as the method does not differentiate between the hardware capabilities of different nodes or the processing required for different requests. All nodes and all requests are treated equally [7].

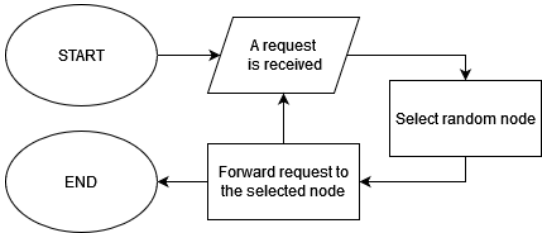


Fig. 3. Flowchart of the random algorithm

For each request received, the *random algorithm* assigns a random node on which it will be processed. The flowchart of this algorithm is shown in Fig. 3. With many nodes, it can perform quite well, as the chance of the same node being selected multiple times in a row diminishes greatly. This also applies vice versa; in smaller-scale solutions, the chance of the same node being selected consecutively increases, leading to possible overloading. As with the previous method, the main advantage is the ease of implementation and scalability [8].

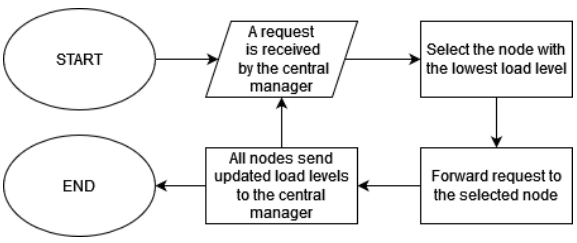


Fig. 4. Flowchart of the central manager algorithm

In the *central manager method*, a central processor maintains information about each node's load and, for each request, selects the node with the lowest load level. When a node's load level changes, it forwards this information to the central manager. The flowchart of this algorithm is shown in Fig. 4. This requires a large amount of inter-node communication, which can create bottlenecks, especially when the load level changes often. Another bottleneck can be created at the central node when multiple requests arrive simultaneously, making choosing an appropriate node difficult [9].

Multiple versions of the *thresholding algorithm* exist, but we have focused on the two we saw as relevant to our paper. In the centralized version, as described in more detail by Chowdhury et al. [10], the nodes are split into two groups - underloaded and overloaded. This can be based on the number of requests processed or CPU load level. This information is forwarded to the central node, which keeps track of both lists. Whenever a request arrives, a random underloaded node is selected, and the request is forwarded to it. In the decentralized version, as described in more detail by Garala et al. [11], each node keeps track of other nodes' load levels. Whenever a node passes through a threshold, it updates the rest of the nodes. When a request arrives, it is processed locally if the local node is underloaded. If the local node is overloaded, it selects a random underloaded node and forwards the request to it. Both version work similarly to the *central manager method* but load levels are updated only when crossing a threshold.

The *central queue* works on the principle FIFO (first in, first out) for both the activity requests - the data to be processed, and for node requests - nodes asking for work. When a node finishes its activity, it asks the central node for the next activity. All the activity requests are stored in a queue on the central node. This can lead to a bottleneck with activities having to wait a long time for a processing node to become available. As this process is harder to visualize in a flowchart, refer to [12] for more details.

V. Intelligent load balancing methods

Using artificial intelligence (AI), the efficiency of traditional load balancing methods can be increased, or new methods might arise. As the nodes can be interpreted as members of a swarm, swarm intelligence represents the most often used field of AI. In this section, we will describe some of the AI-based methods.

A. Ant colony optimization

In this algorithm, the nodes represent food sources that the ants are trying to find. Ants use pheromones to mark their paths, which they take when traveling between nodes, with pheromones increasing as more ants take the same path. We can select the optimal route by assuming that most ants take the best path, concerning the resources available and the distance traveled. This method proves its usefulness in large solutions, with star-bus topology [13].

B. Artificial bee colony optimization

In this method, the "bees" are divided into three groups:

1) employed - workers looking for a food source,
2) onlookers - workers following the employed bees to the best food source,
3) scouts - workers whose food source had been abandoned.

The algorithm starts by dividing all bees into two groups - employed and onlookers. Each employed bee finds a food source and returns to the onlookers, presenting information about the abundance or distance to the food. The onlooker bees then select an employed bee to follow to the source and begin looking for other sources nearby, hoping to find a better solution. If a bee's source is not selected, it becomes a scout bee, looking for a new food source independently [14].

C. Green computing using a neural network

The authors in [15] propose using a neural network (NN) to reduce the energy required for data processing in cloud environments. The NN receives information about the current load levels of all nodes and transforms them into balanced output values. During low load scenarios, the NN recommends shutting down some of the nodes, as they are not required, and the cost of putting a greater load on the rest of the nodes is less than the cost of running a greater number of nodes [16].

D. Combination of multiple methods

Using only one method to solve a complex problem, such as load balancing, may prove difficult. Sui et al. in [17] therefore split the problem into multiple stages and selected a fitting method to solve them separately.

To prevent a node from being overloaded while another node is underloaded, a Support Vector Regression based algorithm was devised. A genetic algorithm was used to select the optimal parameters. By using this approach, the authors were able to predict the future load of the node accurately.

The results from this algorithm were then used in a k-means clustering algorithm to find the nodes to which the workload might be transferred. The nodes were divided into twelve groups based on load level, migration cost, and performance interference. After selecting the appropriate nodes, another algorithm was used to select the best one.

By using an adaptive differential evolutionary algorithm, the authors minimized the network traffic, migration cost, energy consumption, and performance interference of the migration. The result of this algorithm was the node to which the workload should be transferred. The authors succeeded in greatly reducing the network traffic between nodes, as well as energy requirements.

VI. Computation offloading

The concept of computation offloading is an important part of the IoT, as the devices are often constrained by their small size, which only allows for limited processing power. By utilizing resources available elsewhere, on the edge, or in the cloud, more complex tasks can be solved without the user knowing about them. The ideal solution in terms of raw computing performance is the usage of cloud computing, as the data centers offer the latest and greatest processing units. The issue arises from a distance between the device and the data center. Small devices often offer slower connection speeds, which result in a higher time to transfer, possibly offsetting the advantages gained by faster processing time. The solution to this problem might be offered by the utilization of edge computing. By bringing the processing power closer to

the device itself, the time to transfer decreases dramatically while offering increased processing power.

As proposed by Xu et al. in [18], a combination of thing-edge-cloud processing offers a low latency, high-performance solution. When a device requires a processing workload, it can contact a nearby edge device. If the device has the capacity, it will process this workload. If not, it will respond to the device accordingly. The device can then decide whether to process the workload locally or send it to the cloud.

Latency and processing power are not the only focus of task offloading. Samie et al. in [19] focused on using the gateway bandwidth. By swapping between the amount of processing being offloaded, they achieved longer battery life on their device, which is crucial for use in healthcare.

Another algorithm for dynamic offloading was described by Shahhosseini et al. in [20]. By continuously measuring the response times of different nodes, the authors achieved better average response times than any singular node.

VII. Future work

One of the main advantages of edge computing is the decrease in latency compared to a purely cloud-based solution. Traditional methods used in distributed computing have their drawbacks. Load balancing focuses on balancing load across all nodes to prevent overloading or underloading. Computation offloading requires communication between the device and an edge node whenever the processing is required, leading to an increase in latency even when data needs to be processed quickly.

In our diploma thesis, we achieved lower response times as we increased the frequency of our requests. As the CPU load increased, it tried to compensate by increasing its frequency, leading to faster processing times. We want to investigate further whether this can be used to gain an advantage over existing solutions.

By measuring the load added by the data processing on different devices, we can create an artificial intelligence model to predict the load whenever new data arrives.

By maintaining a constant overview of the load levels and CPU frequencies of each node in the solution and predicting the processing load required for a certain workload, we can select the node which will be able to process it without overloading while still offering the highest speed achievable by the CPU. We will then compare whether the gain from processing on a weaker node currently running at the highest possible frequency will offset the transportation time and processing speed achievable by the stronger nodes located further from the device.

Using containerization, we can run multiple data processing applications on the same device, whether on the edge or in the cloud. This should also prove useful in case some applications require more processing power than the edge device can provide, in which case we can deploy the same application to the cloud.

As our research group already possesses a wide array of devices for use in edge computing and several high-power servers, which can serve as the fog layer in our solution, we can test many combinations and conditions under which the solution can operate. The testing will be done using a custom synthetic data generator we have created as a part of our diploma thesis.

Acknowledgment

This publication was supported by the APVV grant ENISaC - Edge-eNabled Intelligent Sensing and Computing (APVV-20-0247) and FEI grant iLan - Indoor localization and navigation (FEI-2023-96).

References

[1] M. De Donno, K. Tange, and N. Dragoni, "Foundations and evolution of modern computing paradigms: Cloud, iot, edge, and fog," *IEEE Access*, vol. 7, p. 150936–150948, Oct 2019.

[2] A. A. Laghari, K. Wu, R. A. Laghari, M. Ali, and A. A. Khan, "A review and state of art of internet of things (iot)," *Archives of Computational Methods in Engineering*, vol. 29, no. 3, p. 1395–1413, Jul 2021.

[3] A. Rashid and A. Chaturvedi, "Cloud computing characteristics and services a brief review," *International Journal of Computer Sciences and Engineering*, vol. 7, no. 2, p. 421–426, Feb 2019.

[4] A. Khalifeh, F. Mazunga, A. Nechibvute, and B. M. Nyambo, "Microcontroller unit-based wireless sensor network nodes: A review," *Sensors*, vol. 22, no. 22, p. 8937, Nov 2022.

[5] D. Fernández-Cerero, J. Y. Fernández-Rodríguez, J. A. Álvarez García, L. M. Soria-Morillo, and A. Fernández-Montes, "Single-board-computer clusters for cloudlet computing in internet of things," *Sensors*, vol. 19, no. 13, p. 3026, Jul 2019.

[6] P. Dukan, A. Kovari, and J. Katona, "Low consumption and high performance intel, amd and arm based mini pcs," *2014 IEEE 15th International Symposium on Computational Intelligence and Informatics (CINTI)*, Nov 2014.

[7] M. E. Mustafa, "Load balancing algorithms round-robin (rr), leastconnection, and least loaded efficiency," *GESJ: Computer Science and Telecommunications 2017|No.1(51)*, 2017.

[8] A. E. Khedr and A. M. Idrees, "Adapting load balancing techniques for improving the performance of e-learning educational process," *Journal of Computers*, p. 250–257, Mar 2016.

[9] P. Beniwal and A. Garg, "A comparative study of static and dynamic load balancing algorithms," *International Journal of Advance Research in Computer Science and Management Studies*, vol. 2, no. 12, p. 386–392, Dec 2014.

[10] S. Chowdhury and A. Katangur, "Threshold based load balancing algorithm in cloud computing," *2022 IEEE International Conference on Joint Cloud Computing (JCC)*, 2022.

[11] K. Garala, N. Goswami, and P. D. Maheta, "A performance analysis of load balancing algorithms in cloud environment," *2015 International Conference on Computer Communication and Informatics (ICCCI)*, Aug 2015.

[12] S. Kaur and G. Kaur, "A review of load balancing strategies for distributed systems," *International Journal of Computer Applications*, vol. 121, no. 18, 2015.

[13] K. Nishant, P. Sharma, V. Krishna, C. Gupta, K. P. Singh, N. Rastogi, and R. Rastogi, "Load balancing of nodes in cloud using ant colony optimization," *2012 UKSim 14th International Conference on Computer Modelling and Simulation*, 2012.

[14] S. M. Ghafari, M. Fazeli, A. Patooghy, and L. Rikhtechi, "Bee-mmt: A load balancing method for power consumption management in cloud computing," *2013 Sixth International Conference on Contemporary Computing (IC3)*, Aug 2013.

[15] N. M. A. Sallami, A. A. daoud, and S. A. A. Alousi, "Load balancing with neural network," *International Journal of Advanced Computer Science and Applications*, vol. 4, no. 10, 2013. [Online]. Available: http://dx.doi.org/10.14569/IJACSA.2013.041021

[16] M. Mesbahi and A. Masoud Rahmani, "Load balancing in cloud computing: A state of the art survey," *International Journal of Modern Education and Computer Science*, vol. 8, no. 3, p. 64–78, Mar 2016.

[17] X. Sui, D. Liu, L. Li, H. Wang, and H. Yang, "Virtual machine scheduling strategy based on machine learning algorithms for load balancing," *EURASIP Journal on Wireless Communications and Networking*, Jun 2019.

[18] X. Xu, Q. Liu, Y. Luo, K. Peng, X. Zhang, S. Meng, and L. Qi, "A computation offloading method over big data for iot-enabled cloud-edge computing," *Future Generation Computer Systems*, vol. 95, p. 522–533, Jun 2019.

[19] F. Samie, V. Tsoutsouras, L. Bauer, S. Xydis, D. Soudris, and J. Henkel, "Computation offloading and resource allocation for low-power iot edge devices," *2016 IEEE 3rd World Forum on Internet of Things (WF-IoT)*, Dec 2016.

[20] S. Shahhosseini, A. Anzanpour, I. Azimi, S. Labbaf, D. Seo, S.-S. Lim, P. Liljeberg, N. Dutt, and A. M. Rahmani, "Exploring computation offloading in iot systems," *Information Systems*, vol. 107, p. 101860, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306437921000910

Architectures for Joint Vehicular Networks

[1]*Dušan HERICH (2nd year),*
Supervisor: [2]*Ján VAŠČÁK*

[1,2]Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1]dusan.herich@tuke.sk, [2]jan.vascak@tuke.sk

***Abstract*—In this paper, we investigate the architecture of IoV and IoFV networks, which represent two emerging fields in transportation. We analyze communication, computing challenges and requirements associated with these networks and propose solutions to overcome these challenges. Our results provide insights into the potential impact of these networks on the transportation sector and on society as a whole.**

***Keywords*—architectures, IoV, IoFV, transportation**

I. Introduction

With the steady growth in traffic volume, the need for effective traffic flow management has become critical [1]. Interconnecting individual vehicles is a potential solution that could enhance the effectiveness and safety of current traffic systems [2]. However, new means of transport such as quadrocopters, also known as drones or UAVs, have caused damages to infrastructures, vehicles, and health. Effective cooperation among different systems could improve the planning of routes, collision avoidance, data collection, and infrastructure sharing. Ground and aerial vehicles are constrained in resources for power, computation, and storage capacity [3]. Vehicular networks, including cloud, edge, and fog computing, offer secure and reliable connections and virtually unlimited computing and storage resources. This paper aims to provide an insight to technologies and approaches used in transport systems for ground and aerial vehicles, identifying differences and common grounds to enable easier cooperation and infrastructure sharing. The focus is put on the Internet of Vehicles [4] and the Internet of Flying Vehicles.

II. Initial Status

The previous year's publication presented a comprehensive review of the architectures of vehicular networks, exploring the underlying principles and concepts of layered models, communication, and enabling technologies. Specifically, the study focused on the significant advancements made in the field of vehicular networking, including the integration of mobile ad hoc networks (MANETs) and infrastructure-based networks to form the so-called vehicular ad hoc networks (VANETs) [5].

While VANETs have enabled the development of a wide range of applications, including traffic management, safety services, and entertainment services, they have been plagued by several challenges, including scalability, security, and reliability. To address these issues, the concept of the Internet of Vehicles (IoV) has emerged, which leverages Internet of Things (IoT) technologies to enable secure data sharing, data processing, and intelligent traffic management among objects in the network [6].

Especially, the study by [7] is of particular interest, which proposed an innovative architecture for a network of connected vehicles performing simultaneous localization and mapping (SLAM). This architecture integrates cloud computing for the efficient merging and storage of created maps, while prioritizing swift horizontal scalability and full automation in the network. By embracing a fully autonomous approach, including automated connection management of new vehicles, network load-balancing, and map creation and storage, the system has demonstrated the potential to enhance the vehicular network space.

III. Objectives Solved in Previous Year

In order to design and test an architecture for joint network we had to create a reliable, separate test-beds for flying and ground vehicles. In the case of ground vehicles we focused on deploying a vehicle-(edge)-cloud [8] system which would allow us to design and test all layers of the IoV as described in earlier sections. Our initial studies were focused on development of intelligent offloading service [9] in line with the Cognitive IoV (CIoV) as shown in the Figure 1.

The service employed machine learning to determine scenarios where offloading a path planning from edge to the cloud would be beneficial. We have measured differences in the time the system requires to produce a solution, that is, to create a path plan from the moment a request for a path is received. This period, therefore, includes the time required to make an offloading decision, time to potentially offload the problem, as well as path plan generation itself [10]. In order to perform the measurement, we have set up four scenarios with maps for path planning ranging from 10x10 cells to 500x500 cells:

- Exclusive edge utilization,
- Exclusive cloud utilization,
- Offloading mechanism with KNN classifier,
- Offloading mechanism with an LR classifier.

Utilizing this system has shown that when the map side size did not exceed 50 cells, it was up to 25.22% faster at 10 cells when the computation was performed only at the edge, and the offloading mechanism was not employed. Consequently, with large maps of a side size of at least 400 cells, the scenario when the only cloud was employed showed a negative time difference up to 6.72%; therefore, employment of only the cloud would yield a path plan faster. In contrast, utilization of the offloading mechanism has shown up to 93.74% saved time with small maps compared to cloud-only deployment and 11.76% in comparison with edge-only employment. Therefore, differences in solution time are smaller with large maps compared to differences in path plan calculation due to the response time from the cloud.

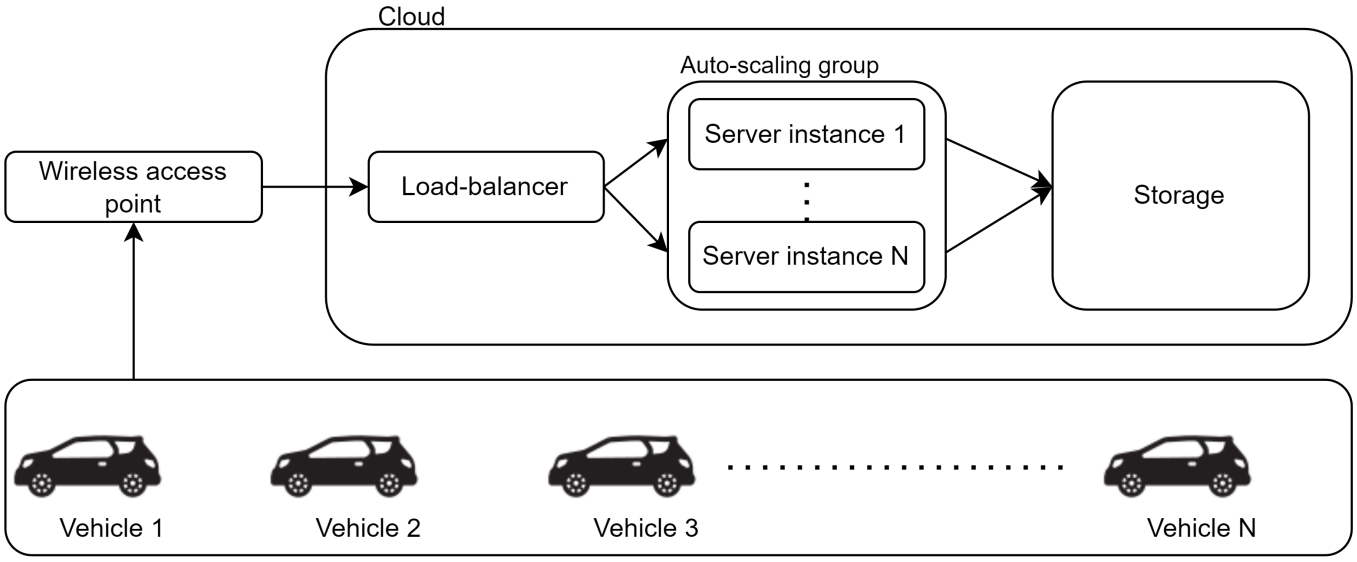


Fig. 1. Architecture of vehicular network.

Currently, we are undertaking a similar approach with flying vehicles. However, because those have limited onboard computing capacity and therefore inability to utilize existing frameworks, it is necessary to create a proprietary framework which will facilitate experimentation simmilar to the one we have performed with ground vehicles.

For this reason, the focus is currently put on creating a system able to manage multiple connected quadrocopters, establish a connection [11] to the base stations and finally to connect them to a cloud. In addition, a functionality to enable remote over-the-internet control and data acquisition to cloud is being created.

IV. Future Work

As a future work, our research aims to propose a referential architecture that will enable effective collaboration between wheeled and flying vehicles in vehicular networks. The proposed architecture will address the unique challenges associated with the integration of different types of vehicles in the network and provide a framework for the secure and efficient exchange of information between them. The proposed architecture will also serve as a basis for further research and development in this exciting and rapidly evolving field.

V. Conclusion

IoV and the Internet of Flying Vehicles (IoFV) are significant areas of research and development in vehicular networking, with the potential to transform the transportation industry by enabling secure, efficient, and intelligent interactions between vehicles and their surroundings. Advanced technologies such as cognitive radio to use the best communication chanel and avoid disruptions in connection, and edge/fog computing are necessary for the development of these networks. It is important to note the unique challenges and requirements associated with communication and computing in a flying environment for IoFV. Continued research and development in these areas are essential for realizing the full potential of IoV and IoFV and shaping the future of the transportation sector.

Acknowledgment

This publication was supported by the VEGA grant EDEN: EDge-Enabled intelligeNt systems (1/0480/22).

References

[1] K. Raghavan, K. J. Ooi, Q. Y. Tan, M. A. Bhuiyan, B. V. Kumar, C. W. Yuen, and M. B. Reaz, "Smart traffic systems guided by principles of traffic circuit theorems," in *2020 IEEE 8th R10 Humanitarian Technology Conference (R10-HTC)*. IEEE, 2020, pp. 1–5.

[2] M. Y. Arafat and S. Moh, "Routing protocols for unmanned aerial vehicle networks: A survey," *IEEE access*, vol. 7, pp. 99 694–99 720, 2019.

[3] S. Jamil and M. Rahman, "A comprehensive survey of digital twins and federated learning for industrial internet of things (iiot), internet of vehicles (iov) and internet of drones (iod)," *Applied System Innovation*, vol. 5, no. 3, p. 56, 2022.

[4] "SCYR 2022: 22nd Scientific Conference of Young Researchers." Faculty of Electrical Engineering and Informatics Technical University of Košice, 2022. [Online]. Available: http://scyr.kpi.fei.tuke.sk/wp-content/scyr-files/history/SCYR_2022_Proceedings.pdf

[5] S. Dalal, B. Seth, V. Jaglan, M. Malik, N. Dahiya, U. Rani, D.-N. Le, and Y.-C. Hu, "An adaptive traffic routing approach toward load balancing and congestion control in cloud–manet ad hoc networks," *Soft Computing*, vol. 26, no. 11, pp. 5377–5388, 2022.

[6] F. Arena and G. Pau, "An overview of vehicular communications," *Future Internet*, vol. 11, no. 2, p. 27, 2019.

[7] D. Herich and J. Vaščák, "Multi-vehicle slam in iov networks," in *2022 Cybernetics & Informatics (K&I)*. IEEE, 2022, pp. 1–6.

[8] A. A. Nasir, "Latency optimization of uav-enabled mec system for virtual reality applications under rician fading channels," *IEEE Wireless Communications Letters*, vol. 10, no. 8, pp. 1633–1637, 2021.

[9] J. Qiu, Y. Chen, X. Zhang, Q. Liu, W. Li, Y. Pei, and L. Liu, "Standardization evolution and typical solutions of iov," in *2019 28th Wireless and Optical Communications Conference (WOCC)*. IEEE, 2019, pp. 1–4.

[10] Y. Zhou, C. Pan, P. L. Yeoh, K. Wang, M. Elkashlan, B. Vucetic, and Y. Li, "Communication-and-computing latency minimization for uav-enabled virtual reality delivery systems," *IEEE Transactions on Communications*, vol. 69, no. 3, pp. 1723–1735, 2020.

[11] A. Arooj, M. S. Farooq, T. Umer, and R. U. Shan, "Cognitive internet of vehicles and disaster management: a proposed architecture and future direction," *Transactions on Emerging Telecommunications Technologies*, vol. 33, no. 8, p. e3625, 2022.

Clustering UI test cases for effective prioritization and selection

[1]*Filip GURBÁĽ (3rd year),*
Supervisor: [2]*Jaroslav PORUBÄN*

[1,2]Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

[1]filip.gurbal@tuke.sk, [2]jaroslav.poruban@tuke.sk

***Abstract*—Specification-based testing can introduce redundancies in the test cases from an implementation perspective. To reduce these redundancies the tester needs to learn the implementation domain. In this paper, we present the method to easily show the tester the impact of their test cases on the implementation using code coverage metrics. Besides the information about the test case uniqueness, our method also helps with test prioritization and selection thanks to hierarchical clustering based on test case similarities.**

***Keywords*—code coverage, hierarchical clustering, UI testing**

I. Introduction

The estimations say that up to 80% of the software development cost is spent on detecting and fixing faults [1]. We want to support software testing and help testers to better understand the software implementation and encourage discussion with developers, so they make less effort in testing software and gain better results.

In many cases, they are UI testers that know only the application domain and test only against the software specification. However, this approach will produce test cases that may be redundant from the implementation perspective. For example, when the UI application is assembled from reusable components the tester may design test cases for every component just because they are in different locations.

Another challenge for the tester is to optimally select test cases to retest when the application is updated. A simple way is to just choose test cases that verify features important to the customers. But the important feature may not be the one that will effectively verify the biggest part of the implementation.

In this paper, we apply our method from previous research to help the UI tester better understand the impact of test cases on the implementation. Our method provides two outputs: a dendrogram with hierarchically clustered test cases useful for test case prioritization and selection, and code coverage difference value between test cases to inform the tester of the amount of redundancy.

II. Related Work

Software testing is researched for a long time during which a few key topics arise. Three probably the most resonating problems in software testing are surveyed by Yoo and Harman: test reduction, test selection, and test prioritization [2].

Code coverage criteria are still commonly used for any of the mentioned problems. The same authors used code coverage as one of three criteria for test case reduction [3]. For the test selection problem, Rothermel and Harrold [4] created a framework to automatically run regression test cases that are affected by the change in the source code.

Huge research is done regarding test prioritization, as surveyed by Mukherjee, Rajendrani, and Patnaik [5]. There are a lot of ways how to approach prioritization using various tools.

Srikanth and Barenjee [6] selected the simple method of prioritizing UI test cases based on a combination of criteria. It is similar to using only code coverage as criteria, but for UI and system testing it is not very precise, so authors are looking for additional criteria for evaluating test cases. The combination that Srikanth and Barenjee used is the customer-assigned priority, developer-perceived implementation complexity, fault proneness, and requirements volatility. However, these criteria need to be assigned manually to test cases before prioritization can be done.

Nguyen and Le [7] used reinforcement learning to prioritize automated UI tests. They defined a weighted coverage graph to model the underlying association among test cases for UI testing.

Yoo, Harman, Tonella, and Susi [8] tried hierarchical clustering for test prioritization. Instead of code coverage they used dynamic runtime behavior (represented as a string) as criteria. They claim that their method is more effective than coverage-based prioritization in terms of the rate of fault detection.

In this paper, we use simple hierarchical clustering based on code coverage, but our goal is to create a tool for the tester to help him understand some aspects of the implementation domain and motivate them to open discussion with developers.

III. The Method

In [9] we proposed a method to improve processes of software testing using code coverage analysis. The goal was to use metadata from program execution during testing to evaluate the complexity of test cases and to automate test case categorization. The method consisted of three main challenges:

A. Design a data structure for code execution tracking,
B. find metrics for measuring test case complexity, and
C. propose a method for clustering test cases based on execution similarities.

We continued to solve these challenges and we applied the method on a commercial project with automated test cases. The illustration of the original method was adjusted to better match our application and the adjusted figure 1 already contains our results. As can be seen, test cases are named

HLx, where X is the number identificator within the project, and the percentage value represents function coverage.

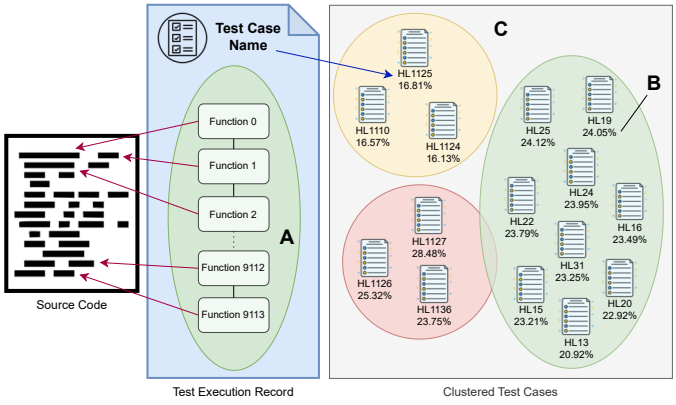


Fig. 1. Use of test execution records for test case clustering by their similarity. **A -** illustration of execution record (function coverage data); **B -** function coverage value of corresponding test case; **C -** similar test cases clustered to three clusters.

For the execution tracking (challenge **A.**) we used code coverage analysis gathered by the commonly used NYC tool[1]. The execution record is a list of all functions in the source code of the user interface application with the information if they were called during execution or not.

As part of challenge **B.** we decided to use code coverage metrics from execution records to determine the test case "complexity". Since we are focusing on the UI tester who is not familiar with the source code, the absolute value of the coverage is not very informational. We will be interested in the difference between two or more coverage metrics so we can inform the tester how much is the test case different from others. Therefore we could use any of the available metrics (line, function, statement, and branch coverage metric), so we decided to use function coverage considering the amount of data needed to be processed and the size of coverage value.

We used hierarchical clustering to cluster test cases by their similarity (challenge **C.**). From a Euclidean distance, Hamming distance, and intersection over union, the last one rendered the best results. We visualized clustering with dendrogram in 2. The linkage between clusters was performed with the average method.

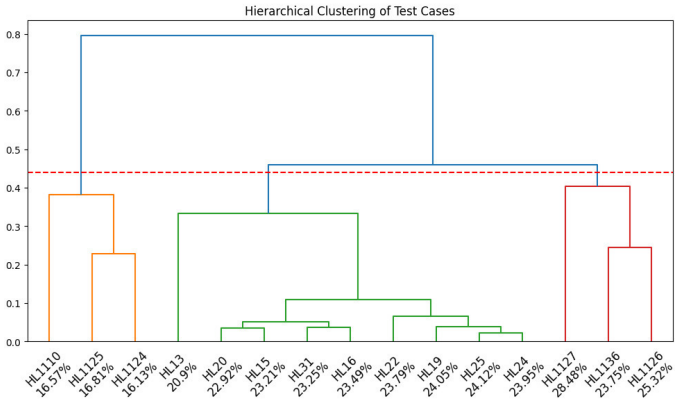


Fig. 2. Dendrogram of hierarchical clustering of test cases. **Distance** is measured by intersection over union; **linkage** is performed by average method; **threshold** was determined visually to mark three clusters.

[1] https://github.com/istanbuljs/nyc

The dendrogram allowed us to determine the distance threshold to form three main clusters. Every cluster represents a feature in the project, so the results are not very surprising.

With this tool, we aim to help the UI tester to better understand the implementation domain and consider it in their estimations. UI testers may have a limited amount of time to perform (or update) all test cases, so they need to select the most efficient ones and also start from the top. In our case, if the tester is given the time for only three test cases, based on the dendrogram they can choose one test case from every cluster with the highest coverage.

Dendrogram is not the only output of our work in this paper. As we mentioned above, the UI tester is interested in the difference between two or more test cases. This is especially helpful to support the tester's understanding of the implementation domain even when they have no knowledge about the source code. After creating the new test case the tester can see how much of it is unique from others in the percentage value (same as coverage value). We encourage testers to open discussions with the developers when they did not expect the amount of coverage to increase after creating a new test case. For example, information like "every form in the UI uses the same code" can change the tester's perspective on the application.

Acknowledgment

This work was supported by project VEGA No. 1/0630/22 "Lowering Programmers' Cognitive Load Using Context-Dependent Dialogs"

References

[1] M. Hossain, "Challenges of software quality assurance and testing," *International Journal of Software Engineering and Computer Systems*, vol. 4, no. 1, pp. 133–144, 2018.

[2] S. Yoo and M. Harman, "Regression testing minimization, selection and prioritization: a survey," *Software testing, verification and reliability*, vol. 22, no. 2, pp. 67–120, 2012.

[3] ——, "Pareto efficient multi-objective test case selection," in *Proceedings of the 2007 international symposium on Software testing and analysis*, 2007, pp. 140–150.

[4] G. Rothermel and M. J. Harrold, "A framework for evaluating regression test selection techniques," in *Proceedings of 16th International Conference on Software Engineering*. IEEE, 1994, pp. 201–210.

[5] R. Mukherjee and K. S. Patnaik, "A survey on different approaches for software test case prioritization," *Journal of King Saud University-Computer and Information Sciences*, vol. 33, no. 9, pp. 1041–1054, 2021.

[6] H. Srikanth and S. Banerjee, "Improving test efficiency through system test prioritization," *Journal of Systems and Software*, vol. 85, no. 5, pp. 1176–1187, 2012.

[7] V. Nguyen and B. Le, "Rltcp: A reinforcement learning approach to prioritizing automated user interface tests," *Information and Software Technology*, vol. 136, p. 106574, 2021.

[8] S. Yoo, M. Harman, P. Tonella, and A. Susi, "Clustering test cases to achieve effective and scalable prioritisation incorporating expert knowledge," in *Proceedings of the eighteenth international symposium on Software testing and analysis*, 2009, pp. 201–212.

[9] F. Gurbál' and J. Porubän, "Using coverage metrics to improve system testing processes," *SCYR 2022: Nonconference Proceedings of Young Researchers*, pp. 215–216, 2022.

Lung ultrasound sign classification

[1]*Maroš HLIBOKÝ (2nd year),*
Supervisor: [2]*Marek BUNDZEL*

[1,2]Dept. of Cybernetic and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1]maros.hliboky@tuke.sk, [2]marek.bundzel@tuke.sk

***Abstract*—The identification of lung artifacts through the use of lung ultrasound aids in the diagnosis of various conditions, and current research aims to automate the detection of such artifacts via machine learning. Our proposed solution utilizes analytical computer vision techniques to detect two specific types of lung artifacts, namely A- and B-lines.**

***Keywords*—classification, image processing, lung ultrasound, pathology detection**

I. Introduction

Lung diseases are a common worldwide issue, making it important to accurately diagnose and monitor them. The COVID-19 pandemic has led to an increased interest in collecting clinical data and utilizing artificial intelligence through machine learning. One method of diagnosis is lung ultrasound (LUS), which is safe, portable, and cost-effective. Unlike other diagnostic techniques like chest X-rays, CT scans, and MRI scans, ultrasound technology does not expose patients to ionizing radiation. Ultrasound devices are also more accessible and cost-effective compared to other methods, allowing practitioners to examine more patients [1].

II. Lung signs

In this paper we propose an analytical approach for the automatic classification of A- and B-lines. A-lines are visible on an ultrasound screen as horizontal lines parallel to the pleural line. They occur when ultrasonic waves reflect back and forth between the pleura and the transducer, creating an echo that is consistently spaced and corresponds to the distance between the skin and the pleura. we can say that A-lines are the first indicator of healthy lungs.

B-lines are vertical artifacts that are clinically defined as starting from a point along the pleural line and extending downwards to the bottom of the ultrasound screen, perpendicular to the pleural line. These lines are created when sound waves enter the pleural cavity, and they move in sync with the patient's inhalation and exhalation [2], [3].

III. Dataset

In our initial research, we utilized the POCUS (point of care ultrasound) dataset, which was created by Born and is freely available. This dataset consists of samples from COVID-19 patients, patients with bacterial pneumonia, viral pneumonia without COVID-19, and healthy patients [4]. It contains 261 recordings, including 202 videos and 59 images taken by convex and linear probes from 216 patients. We focused on recordings that were labeled based on the presence or absence of A- and B-lines, which allowed us to define four classes. We only used videos to capture the dynamic nature of data during an examination and prioritized the possibility of using analytical methods for real-time image assessment.

TABLE I
POCUS dataset overview

no A- and B-lines present (class X)	247
A-lines present (class A)	23
B-lines present (class B)	71
A- and B-lines present (class AB)	11

IV. Research Contribution

The aim of our work was to classify ultrasound videos of lungs depending on the presence of A- and B-lines. We prepared the training data by preprocessing and augmenting them. We then implemented an analytical method using Fourier transformation to detect lines, which we then used to classify videos.

A. Data preprocessing

To enhance A- and B-line detection in images, we implemented various preprocessing techniques, including size reduction, information reduction, noise suppression, and eliminating sudden signal changes. We resized the video frames using bicubic interpolation to 300 x 300, and converted the original three-channel RGB images to a single grayscale channel, reducing the amount of information. To reduce noise, we employed two primary methods: convolution with a Gaussian kernel and Fourier transformation [5].

B. Line detection

Detection of A and B lines in lung USG image data involves the identification of lines, for which several methods exist. Hough transformation is commonly used for line detection, but it is not effective in detecting curved A-lines in our image data, which is acquired using a convex ultrasound probe. Convolutional neural networks are another option, but they are computationally complex and require a significant amount of training data.

Our proposed approach involves searching for local maxima both horizontally and vertically in the image, without requiring training data. This method is computationally efficient and can be used for real-time processing. The vertical search detects horizontal maxima for A-line detection, while the horizontal search detects vertical maxima for B-line detection. Candidate

lines are then connected and further evaluated for artefacts or noise.

However, there are often disconnected pixels and lines present, which are resolved using morphological dilation. The candidate lines are then segregated, and invalid candidates are removed by performing contour search and applying a high-pass filter based on the contents of the selected contour.

In the last step of line detection we transformed the detected lines into vectors. Horizontal A-lines were transformed into a constant function (equation 1), vertical B-lines were transformed into a linear function (equation 2), while tangent a and constant y_0 were determined using least square approximation in a way that the resulting line approximated the uneven line as much as possible. Each detected line was then saved as two points (starting and ending) with x and y coordinates on the image. We recorded three pieces of information for both types of lines: the maximum number of lines detected in a single frame throughout the video, the total number of lines detected, and the high-pass filtering result using the length of newly added lines.

$$y_A = y_0 \tag{1}$$

$$y_B = ax + y_0 \tag{2}$$

C. Data analysis

Before video classification, we processed the results of line detection over individual frames by calculating statistical metrics, which is based on accummulated values from buffers.

We defined five metrics that have a significant impact on classification:

- average number of A-lines per frame:

$$\overline{A} = \frac{\sum_{t=1}^{n} A_{buf}(t)}{n}, \tag{3}$$

  where n is the number of frames, and $A_{buf}(t)$ is the number of A-lines identified in frame t (as stored in the buffer);
- maximum number of A-lines on a single frame:

$$A_{max} = max(A_{buf}); \tag{4}$$

- average number of B-lines per frame:

$$\overline{B} = \frac{\sum_{t=1}^{n} B_{buf}(t)}{n}; \tag{5}$$

- maximum number of B-lines on a single frame:

$$B_{max} = max(B_{buf}); \tag{6}$$

- the ration of A- and B-lines:

$$\overline{AB}_{ratio} = \frac{\overline{A}}{\overline{A} + \overline{B}}. \tag{7}$$

D. Classification

For video classification we created a configurable set of IF-THEN heuristic rules based on results over selected videos from the dataset:

- class A - video showing A-lines,
- class B - video showing B-lines,
- class X - classification not possible, expert evaluation is required.

For classification, the following rules were used:

TABLE II
CONFUSION MATRIX OF VIDEO CLASSIFICATION RESULTS

		predicted		
		A	B	X
true	A	4	0	1
	B	0	9	0
	X	1	2	0

1) if the maximum number of A-lines and B-lines is smaller than 2, then the video is of class X:

$$\text{IF} \quad (A_{max} < 2) \wedge (B_{max} < 2) \quad \text{THEN} \quad x \in X, \tag{8}$$

2) if the ratio of A- and B-lines is bigger than 0.5, the average number of B-lines per frame is smaller than 1.5, and the maximum number of B-lines on a frame is smaller than 3, the video is of class A:

$$\text{IF} \quad (\overline{AB}_{ratio} > 0.5) \wedge (\overline{B} < 1.5) \wedge (B_{max} < 3) \quad \text{THEN} \quad x \in A,$$

3) any other video is considered to belong to class B:

$$\text{IF} \quad (x \notin X) \wedge (x \notin A) \quad \text{THEN} \quad x \in B \tag{9}$$

V. Results

We employed three metrics to evaluate the classification outcome: accuracy, balanced accuracy, and a confusion matrix. We conducted tests on a subset of 17 videos from the dataset, comprising of 5 videos displaying A-lines, 9 videos displaying B-lines, and 3 videos displaying neither. Our experiments yielded an accuracy of 76.471%, and a balanced accuracy of 60%.

VI. Conclusion

A software algorithm was developed to locate A and B lines in ultrasonographic video records. The algorithm was capable of identifying and highlighting these lines within the image. A rule-based classifier was devised, tested, and assessed for the algorithm, allowing for the classification of video recordings into appropriate groups. Further modifications to the configuration file can enhance the accuracy of the classifier and the quality of the output. To boost the classifier's accuracy specifically, new rules can be added to the existing production rule set, which will aid in the alteration of the output classification.

Acknowledgment

This publication was realized with support of the Operational Programme Integrated Infrastructure in frame of the project: Intelligent systems for UAV real-time operation and data processing, code ITMS2014+: 313011V422 and co-financed by the Europen Regional Development Found.

References

[1] S. B. Desai, A. Pareek, and M. P. Lungren, "Deep learning and its role in covid-19 medical imaging," *Intelligence-Based Medicine*, vol. 3, p. 100013, 2020.

[2] D. A. Lichtenstein, "Lung ultrasound in the critically ill," *Annals of intensive care*, vol. 4, no. 1, pp. 1–12, 2014.

[3] L. Gargani and G. Volpicelli, "How i do it: lung ultrasound," *Cardiovascular ultrasound*, vol. 12, no. 1, pp. 1–10, 2014.

[4] J. Born, N. Wiedemann, G. Brändle, C. Buhre, B. Rieck, and K. Borgwardt, "Accelerating covid-19 differential diagnosis with explainable ultrasound image analysis," *arXiv preprint arXiv:2009.06116*, 2020.

[5] J. Wolf, "Redundancy, the discrete fourier transform, and impulse noise cancellation," *IEEE Transactions on Communications*, vol. 31, no. 3, pp. 458–461, 1983.

Multiport Power Converter with Planar Transformer – Simulation Verification

[1]*Adrián MARCINEK (3th year)*
Supervisor: [2]Marek PÁSTOR

[1,2]Dept. of Electrical Engineering and Mechatronics, FEI TU of Košice, Slovak Republic

[1]adrian.marcinek@tuke.sk, [2]marek.pastor@tuke.sk

***Abstract*— The paper describes the simulation verification of a multiport power converter with a planar transformer with windings integrated into a printed circuit board (PCB). The designed planar transformer represents the primary element for transmitting power through a multiport converter port. The ANSYS software was used for the verification design of the planar transformer and verification multiport converter with these transformers. Planar transformers represent magnetic coupling between planar transformer ports.**

Keywords*— *planar transformer, windings integrated into PCB, multiport converter

I. INTRODUCTION

This paper follows up on paper [1], where the multiport power converter control method was presented. We consider using a planar transformer for this multiport power converter, which assumes the necessity of a correct design of a planar transformer. A multiport power converter provides more advantages for interconnecting several sources and loads compared to using multiple DC-DC or DC-AC converters connected to a common DC bus [2-5]. Multiport power converter consists of multiple stages, where one stage, which we can also call „port“, consists of full–bridge converter and a power transformer. Multiport converter topologies can use transformers with standard iron cores or ferrite cores with windings wound from copper wire or copper strip [6-9]. The topology of multiport power converter with five ports is shown in Fig. 1 In our simulation verification, we will use three port multiport converter with equal transformers.

The paper focuses on a multiport power converter with three ports and magnetic coupling between these ports. Magnetic coupling using a connection of secondary windings of planar transformers. The transformer parameters depend on the maximum voltage and current parameters of the full-bridge converter. The paper describes the process of designing the planar transformer and simulation verification of the three-port multiport converter with this transformer.

II. DESIGN OF PLANAR TRANSFORMER

The main difference between traditional and planar transformers is the geometry of the core and winding structure. The core of a planar transformer offers a high value of effective permeability, low profile, low leakage inductance and excellent thermal characteristics. The winding of a planar transformer can be constructed as a stand-alone component, multilayer printed board (PCB) or integrated into the multilayer PCB of the device. The fundamental limitation of windings

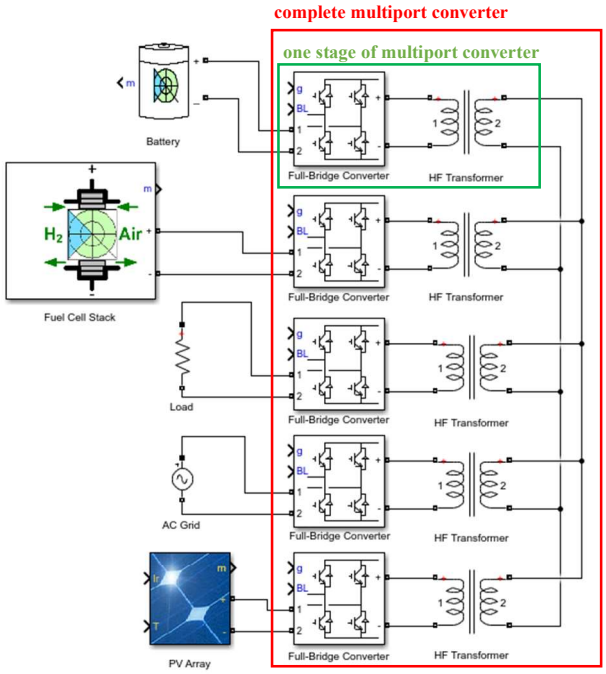


Fig. 1 Topology of multiport power converter

integrated into PCB is maximum current carrying capacity of windings. There are a few default copper foils using for PCB with different thicknesses – 18, 35, 70 and 105 µm. During the design of windings of the planar transformer, it is necessary to consider the distribution of the magnetic flux to the core and the local saturation of the core. Local saturation is possible to prevent using interleaved winding. It means, that the primary, secondary or both windings are divided into several parts or layers connected in serial or parallel. The winding design also influences the leakage inductance, which manifests itself when the transformer is loaded which can cause significant losses. Therefore, the appropriate design of planar transformer windings into individual layers can influence the leakage inductance value.

The [10] describes planar transformer model, where every magnetic layer represents nonlinear inductance, whose magnetic characteristics depend on the hysteresis model depending on temperature. This model can be integrated directly into the simulation software. Paper [11] is more focused on the winding losses calculation and leakage inductance. The paper also describes a comparison of winding losses between the planar ER core and the E core.

A. *Full – bridge converter*

Because our multiport converter consist of three ports, energy flow must be controlled between these ports. It means, that the primary winding of each transformer is connected to a full-bridge converter in our case . The input of this full bridge then represents one port of the multiport converter. This full–bridge converter ensures control of power transferred through individual ports and its parameters are important for planar transformer design. The most important parameters of full – bridge inverter for planar transformer design are:

- maximum input/output voltage: 50 V,
- maximum input/output current: 50 A.

The planar transformer was designed based on the parameters described above.

B. *Transformer Core Selection*

The selection of transformers and their core is important for the selection of a properly control method of the multiport converter. The required planar transformers parameters were selected based on parameters of available converter and are listed in TABLE 1.

TABLE I. REQUIRED PARAMETERS OF TRANSFORMER

Parameter	Value	Meaning
U_1	24 V	amplitude of input voltage
U_2	48 V	amplitude of output voltage
I_{1MAX}	36 A	max. current of primary winding
I_{2MAX}	18 A	max. current of secondary winding
N_1	3	number of primary turns
N_2	6	number of secondary turns
f	100 kHz	supply voltage frequency

Due to core availability, the Ferroxcube E64/10/50 + PLT64/50/5 core with 3C95 core material was selected. TABLE 2 shows the parameters of the selected core. Based on these parameters, further parameters of the transformer were calculated and compared with simulation results.

TABLE II. SELECTED CORE PARAMETERS

Parameter	Value	Meaning
V_e	35500 mm^2	effective volume
S_e	519 mm^2	effective area
l_e	69.7 mm	effective length
A_L	18000	inductance factor

C. *Transformer Winding Design*

The result of the winding calculation determines the required cross-section of wires for the primary and secondary winding. Since windings were integrated into an eight-layer PCB, we could choose only from standard copper foil height which influenced resultant width of the winding copper path. Because there are also dimension limits in the transformer core, we were forced to reinforce both the primary and secondary windings in each layer by another parallel layer. With this reinforcement, we achieved the required current carrying capacity and the required number of turns for each winding. In addition, we also used the method which eliminates local saturation of transformer core–interleaved windings. Interleaved windings means, that the primary windings of the transformer were fitted to the outer layers of the PCB because higher current needs better heat dissipation [12]. Mean lengths of designed winding are:

- primary winding l_1 = 543 mm,
- secondary winding l_2 = 1214 mm.

The designed transformer windings are integrated into the PCB board, where was used the maximum height of standard copper foil – 105 µm. Using this copper foil ensured considering the required current carrying capacity. Fig. 2 shows a 3D model of the designed planar transformer. This 3D model further used in ANSYS simulation software to simulate magnetic induction in transformer core but also for simulation verification of multiport converter.

Magnetic induction in transformer core from ANSYS is in Fig.3 and Fig. 4 shows a real PCB windings and complete planar transformer – ferrite core + PCB windings.

Fig. 2 3D model of designed planar transformer

For additional mechanical protection and increasing voltage strength, a 0,8 mm thick insulating material was inserted above and under the windings. All parameter of designed transformer were later verified by RLC bridge and listed in TABLE 3.

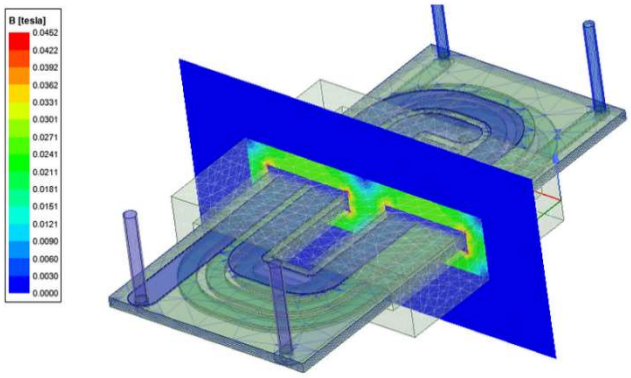


Fig. 3 3D model of planar transformer in ANSYS - simulation of magnetic induction

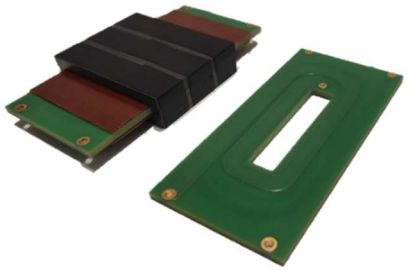


Fig. 4 Complete transformer and windings designed in PCB

TABLE III. COMPARISON BETWEEN CALCULATED AND MEASURED PLANAR TRANSFORMER PARAMETERS

Parameter	Calculated	Measured (1 kHz)	Meaning
B	34.68 mT	24 mT (mean value from ANSYS, 100 kHz)	magnetic induction
L_{1h}	162 µH	168.4 µH	primary inductance
L_{2h}	648 µH	659 µH	secondary inductance
L_{leak}	7.3 nH	-	leakage inductance
R_{prim}	6.76 mΩ	6.8 mΩ	primary resistance
R_{sec}	27.27 mΩ	29.1 mΩ	secondary resistance
I_{μ}	0.37 A	0.37 A (27 um gap – ANSYS)	magnetization current

III. SIMULATION RESULTS

As was mentioned above, the main software for simulation verification of planar transformer in multiport converter was ANSYS. For verification of basic principles of energy flow between ports and simplicity of simulation, we do not use full – bridge converters mentioned in previous chapters. ANSYS is also not adapted for developing different types of control – these types of a simulation take a lot of time. For that, we simplified the simulation only for planar transformers and harmonic power supplies.

Multiport converted was supplied by harmonic voltage with an amplitude of 24 V and frequency of 100 kHz. During the simulation, we sequentially connected individual transformers to the supply voltage or to load and investigated how the transmitted power is distributed without any control method. Fig. 5 shows a simulation model of the multiport converter in ANSYS software. Transformers in the simulation are represented by a 3D model of the designed planar transformer and resistors in the model represent the primary and secondary winding resistance of the transformers. Resistor R7 represented the load on the primary side of the transformer T2 – load on port 2.

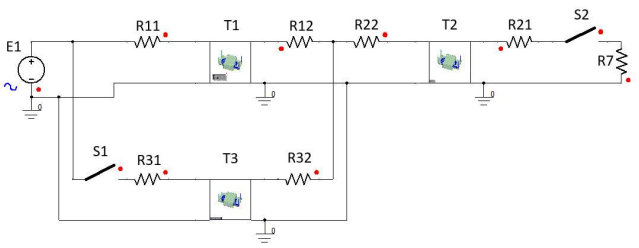


Fig. 5 Multiport converter simulation model with designed planar transformer

The transformers were named T1, T2 and T3 in the simulation. We can see simulation results in Fig. 6, Fig. 7 and Fig. 8. Whereas the simulation procedure was as follows:

- At time $t = 0$ s, a harmonic voltage of amplitude 24 V and frequency 100 kHz was connected to the primary side T1. Through port T1 supplies magnetizing current for all transformers of the multiport converter.
- At time $t = 40$ µs, the same harmonic voltage as on T1 was connected to the primary side of T3. Magnetizing current through T1 decreased and was redistributed between T1 and T3. Voltage in the intermediate circuit remained unchanged.
- At time $t = 80$ µs, a resistive load was connected to the primary winding of T2. T2 started to supply the current 10 A to the load.
- The current supplied to T2 was redistributed between T1 and T3.

IV. CONCLUSION

The paper describes the short procedure of planar transformer design and its simulation verification when used as a main part of the multiport power converter. The planar transformer was selected due to the requirement of high-power density, modularity, and easy and fast assembly. Windings for the planar transformer were selected as integrated into PCB. During the design and verification, ANSYS' extensive capabilities were confirmed. Transformer's 3D model was easy to integrate into ANSYS software for verifying magnetic induction in the transformer core and for verifying the transformers used in multiport converter. The same 3D model was used for both verifications mentioned above. In the simulation experiment, two of three ports of the multiport converter were supplied from the power source and the third was loaded by the resistive load. The result of the simulation showed that the converter can work without any control with harmonic voltages and ideal sources. These results can be incorrect in

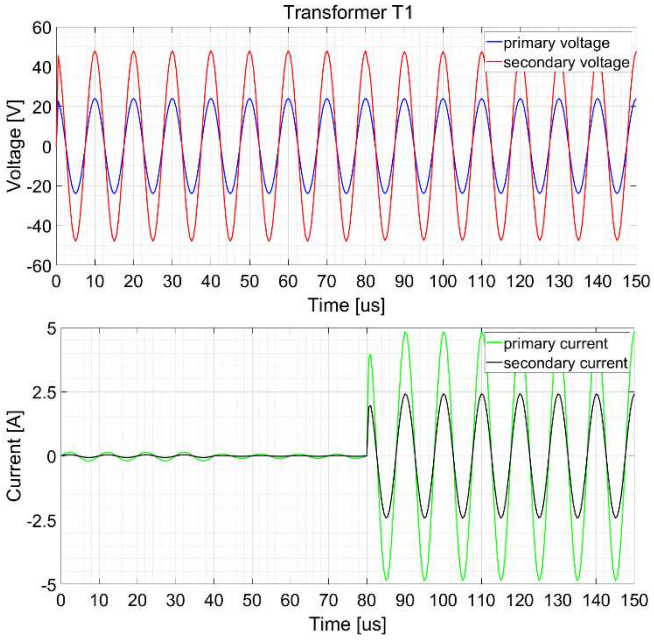


Fig. 6 Primary and secondary voltage and current of T1

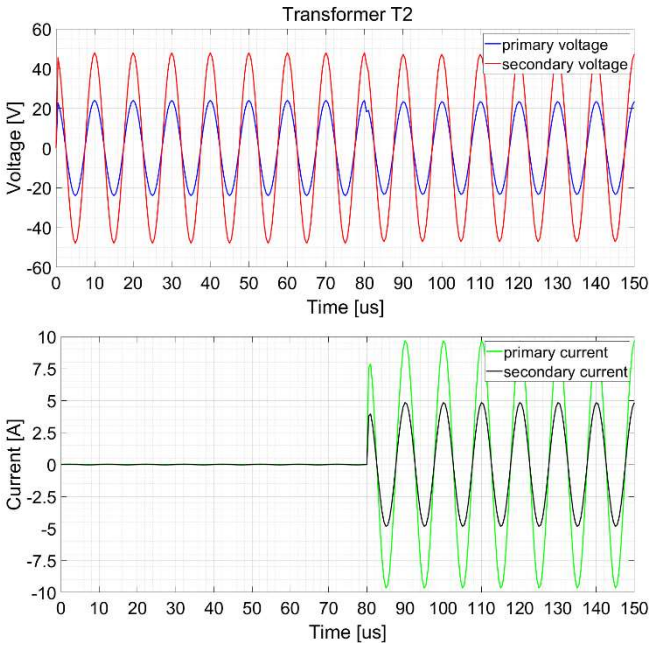


Fig. 7 Primary and secondary voltage and current of T2

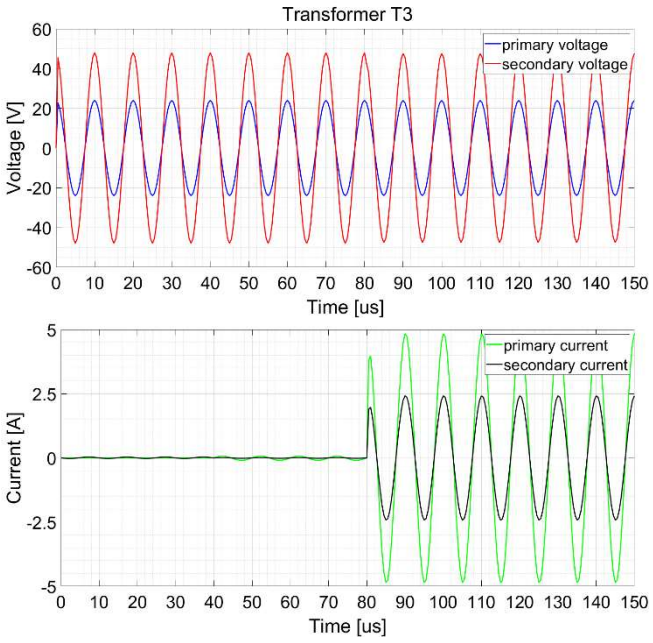


Fig. 8 Primary and secondary voltage and current of T3

real conditions, where energy sources and transformers can have different performance parameters. This fact is possible to solve with a full–bridge converter with proper control.

Further research will be focused on a simulation experiment with rectangle voltages which supplied a multiport converter and later multiport converter powered by DC sources using full–bridge converters and properly control algorithm. Results from these simulations give us a more accurate view of the control methods suitable for efficient energy flow between the port of the multiport converter.

ACKNOWLEDGMENT

This work was supported by the Slovak Research and Development Agency under the contract No. APVV-15-0750.

This work was supported by the Slovak Research and Development Agency under the contract No. APVV-18-0436.

REFERENCES

[1] A. Marcinek, F. Ďurovský, "Control of Multiport Power Converter", In: 22nd Scientific Conference of Young Researchers: proceedings from conference. - Košice (Slovensko): Technická univerzita v Košiciach, s. 152-156 [CD-ROM, print]. - ISBN 978-80-553-4061-6. Available online : http://scyr.kpi.fei.tuke.sk/wp-content/scyr-files/proceedings/SCYR_2022_Proceedings.pdf

[2] Y. M. Chen, A. Q. Huang and X. Yu, "A High Step-Up Three-Port DCDC Converter for Stand-Alone PV/Battery Power Systems," in IEEE Transactions on Power Electronics, vol. 28, no. 11, pp. 5049-5062, Nov. 2013.

[3] Hongfei Wu, Yan Xing, Yanbing Xia and Kai Sun, "A family of non-isolated three-port converters for stand-alone renewable power system," IECON 2011 - 37th Annual Conference of the IEEE Industrial Electronics Society, Melbourne, VIC, 2011, pp. 1030-1035.

[4] L. J. Chien, C. C. Chen, J. F. Chen and Y. P. Hsieh, "Novel Three-Port Converter With High-Voltage Gain," in IEEE Transactions on Power Electronics, vol. 29, no. 9, pp. 4693-4703, Sept. 2014.

[5] B. Wang, M. Sechilariu and F. Locment, "Intelligent DC Microgrid With Smart Grid Communications: Control Strategy Consideration and Design," in IEEE Transactions on Smart Grid, vol. 3, no. 4, pp. 2148-2156, Dec. 2012.

[6] M. McDonough, "Integration of Inductively Coupled Power Transfer and Hybrid Energy Storage System: A Multiport Power Electronics Interface for Battery-Powered Electric Vehicles," in IEEE Transactions on Power Electronics, vol. 30, no. 11, pp. 6423-6433, Nov. 2015.

[7] H. Zhu, D. Zhang, B. Zhang and Z. Zhou, "A Nonisolated Three-Port DCDC Converter and Three-Domain Control Method for PV-Battery Power Systems," in IEEE Transactions on Industrial Electronics, vol. 62, no. 8, pp. 4937-4947, Aug. 2015.

[8] V. N. S. R. Jakka, A. Shukla and G. D. Demetriades, "Dual Transformer Based Asymmetrical Triple-Port Active Bridge (DTATAB) Isolated DC-DC Converter", IEEE Transactions on Industrial Electronics, vol. 64, no. 6, pp. 4549-4560, June 2017.

[9] E. Asa, K. Colak, M. Bojarski and D. Czarkowski, "Asymmetrical Duty-Cycle and Phase-Shift Control of a Novel Multiport CLL Resonant Converter", IEEE Journal of Emerging and Selected Topics in Power Electronics, vol. 3, no. 4, pp. 1122-1131, Dec. 2015.

[10] Lew Andrew R. Tria, Daming Zhang, John E. Fletcher, „ Planar PCB Transformer Model for Circuit Simulation“, School of Electrical Engineering and Telecommunications, University of New South Wales, NSW 2052, Australia, 2015

[11] Riccardo Pittini, Zhe Zhang, Ziwei Ouyang, Michael A. E. Andersen, Ole C. Thomsen, „Analysis of Planar E+I and ER+I Transformers for Low-Voltage High-Current DC/DC Converters with Focus on Winding Losses and Leakage Inductance“, 7th International Power Electronics and Motion Control Conference - ECCE Asia, June 2-5, 2012, Harbin, China, pp. 488 – 493.

[12] Ferroxcube, „ Design of planar power transformers – application note“, online.

Clock generators for Ultra-Wideband Sensor system

[1]*Patrik Jurík (2nd year),*
Supervisor: [2]*Pavol Galajda*

[1,2]Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

[1]patrik.jurik@tuke.sk, [2]pavol.galajda@tuke.sk

***Abstract*—This paper introduces designed and realized High frequency clock generator circuits for Ultra-Wideband (UWB) sensor systems. Implementation of introduced high frequency oscillators are based on dielectric resonator and Application Specific Integrated Circuit voltage controlled oscillator. Two transistor types for oscillators with dielectric resonator were used. Low noise (BJT) bipolar transistor BFR740L3RH and low noise (JFET) junction gate field-effect transistor CE3520K3. Many versions of dielectric oscillators were produced on frequency around 21.75 GHz and 10.85 GHz. The article describes the next research focus and new ideas for the miniaturization.**

***Keywords*—Oscillators, RF, DRO, VCO, UWB, PLL**

I. Introduction

Ultra-Wideband system sensors come to the fore in last two decades, miniaturization and cheaper mass production bring system close to users. Our research focused on developing and designing the Ultra-Wideband sensor based on M-sequence. The UWB sensor consists of several circuits that work together in one system. Few sub circuits were developed especially as Application Specific Integrated Circuits (ASIC) like a wideband amplifiers [1], analog-digital converters [2], UWB transceivers [3], UWB mixers, UWB divider [4]. The UWB sensor can be used in various applications for example in person localizations [5], monitoring farm animals, food products, monitoring people health or non-desctrucive testing. Nowadays still new application come to the fore for example water quality control [6], or health monitoring [7].

The next step of the research was to develop an oscillator circuit for the main signal clock generator of the UWB sensor. The main system clock is a very important and delicate circuit, the performance of the entire system is directly dependent on the clock signal generated by the clock generator. Therefore, it is necessary to pay attention to the parameters of the clock generator, especially the magnitude of the phase noise, frequency stability, signal amplitude, the possibility of tuning the frequency, low power consumption, small size and others. Based on these properties, several oscillators suitable for high-frequency applications operating at frequencies of 21.75 GHz or 10.85 GHz were investigated.

II. Design of oscillator with dielectric resonator

Oscillators with dielectric resonators provide very low phase noise, low energy consumption and provide short-term stable frequency, which met our selection conditions. For designing a dielectric oscillator, several parts are needed: an active device, a dielectric resonator, a printed circuit board (PCB) intended for high-frequency applications, a metal box for shielding the device and a tuning element. The oscillators were designed based on a dielectric element, namely a dielectric puck made of MgTiO3 - CaTiO3 solid solution by T-ceram [8]. The resonator was customized for frequencies of 21.75 GHz and 10.85 GHz.

Two different types of transistors were used in designs, low-noise bipolar transistor BFR740L3RH Infineon [9] and low-noise junction gate field-effect transistor CE3520K3 [10]. Both types of transistors have different pin layouts, so the PCBs had to be optimized for each type. Dielectric resonators were manufactured, trimmed, and measured in a measurement chamber for our purpose by T-ceram. When the manufacturer trimmed the resonator, there may be deviations from the resonant frequency, this technique is an irreversible phenomenon, so it is necessary to take this into account. However, there are tuning techniques that can influence the field of the dielectric resonator and enable the tuning of the resonant frequency of the resonator. Two main principles are mechanical tuning with a microwave tuning element, or electrical tuning with a varactor diode connected to a micro strip line coupled to resonator [11].

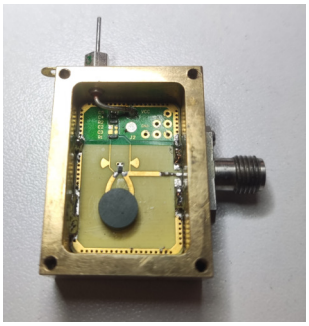
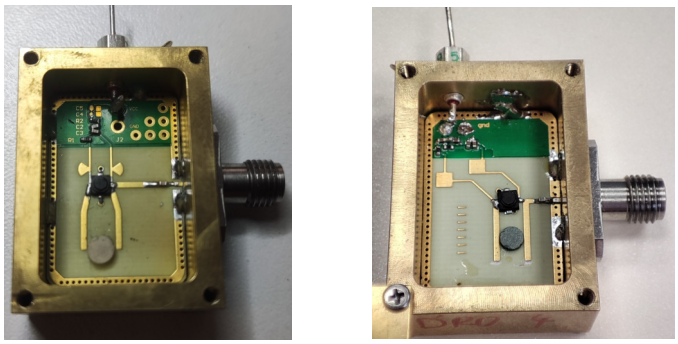


Fig. 1: DRO with BJT transistor BFR740L3RH oscillating at 10.7 GHz.

For designing high-frequency circuits a different approach has to be used, it is necessary to think about PCB parameters, parameters of resonator, transistor and power supply circuits. For circuit simulation a few software can be used, in our case, PathWave Advanced Design System (ADS) and CST STUDIO SUITE were used. Both software offers different methods of simulations. For our oscillators, we used ceramic substrates from Rogers company especially RO4360G2 and RO4003 with dielectric constants 6.15, and 3.38, and thickness 508 [µm]. Many types of oscillators topologies were designed on both substrates with both types of transistors, also few different types of RF transistors were tested. Fig.1 shows an oscillator with dielectric resonator in parallel feedback configuration

with BJT transistor BFR740L3RH. Oscillator frequency is 10.7 GHz. Fig. 2a and 2b present different topologies of the oscillator with dielectric resonator, with a different type of dielectric resonator, which causes a different oscillating frequency 21.75 GHz and 21.68 GHz.



(a) (b)

Fig. 2: DRO with JFET transistor CE3520K3, oscillating frequency around 21.75 GHz (a.). Another topology of DRO with JFET transistor CE3520K3, oscillating frequency is 21.68 GHz, oscillators are based on the different dielectric resonators (b.).

Proposed oscillators have single-ended output, but UWB system also supports differential signal distribution. To fill these conditions new topologies of oscillators have to be discussed. The advantages of dielectric resonators have influenced the design of a new differential oscillator with a dielectric resonator. The oscillator is implemented as an ASIC in high-performance technology 130 nm SiGe BiCMOS with npn-HBTs with cutoff frequency up to f_T = 350 GHz and f_{max} = 450 GHz. The advantage of designing the oscillator as an ASIC lies mainly in the possibility of a symmetric design structure and minimizing the overall circuit into a single package. Which will then be soldered to PCB with microstrip lines and coupled dielectric resonator. We have encountered several problems in properly matching micro strip circuits to ASIC, caused by the influence of parasitic capacitances between pads and inductance on bond-wires. More simulations have to be performed to get better circuit performance on PCB.

III. ASIC Voltage controlled oscillator

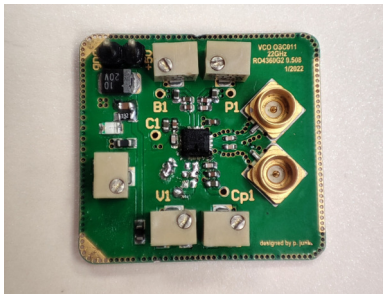


Fig. 3: Evaluation PCB for designed ASIC VCO with bias voltage adjustment.

Fig.3 shows an ASIC circuit consisting of two voltage-controlled oscillators that has been designed and tested in the past [12]. For measurement purposes, a test circuit board with trimmers was fabricated to set 5 different bias voltages. The oscillator generates a signal in the range of 21 GHz to 24 GHz based bias voltages. However, the oscillator does not achieve the required properties mainly the high phase noise amplitude. The oscillator has been tested with a commercially available phase-locked loop circuit ADF4002 [13]. The measured results showed reduced phase noise of the generated signal. The next step is to integrate both circuits on one PCB. Due to the elimination of the high-frequency SMP connectors and coaxial cables, we expect even lower phase noise.

Conclusion

This paper describes the designed oscillators with dielectric resonators and voltage controlled oscillator. Measurement of phase noise and other oscillator parameters was performed with collaboration at VUT BRNO because they have at their disposal measurement equipment with a range of up to 50 GHz. In the near term, we want to test mechanical techniques for tuning the output frequency by moving a metal piston over a dielectric resonator. The next step will be to implement the ASIC DRO circuit into a System-in-Package. Several versions of the SiP with an external clock signal generator have been created to date, so as part of the miniaturization of the device we need to implement a naked die with an oscillator on the SiP substrate.

Acknowledgment

This work was supported by the Slovak Research and Development Agency under the Contract No. APVV-18-0373 and Scientific Grant Agency (VEGA) under the Contract No. 1/0584/20, and by the Faculty of Electrical Engineering and Informatics, TU Košice under the Grant FEI-2023-95.

References

[1] M. Sokol, P. Galajda, S. Slovak, and M. Pecovsky, "Differential amplifier based on cherry-hooper topology for uwb m-sequence radars," in *2019 IEEE Radar Conference (RadarConf)*. IEEE, 2019, pp. 1–5.

[2] M. Sokol and P. Galajda, "Design of a 7-bit flash ad converter for m-sequence uwb sensor systems," in *2021 31st International Conference Radioelektronika (RADIOELEKTRONIKA)*. IEEE, 2021, pp. 1–6.

[3] P. Galajda, S. Slovak, M. Sokol, M. Pecovsky, and M. Kmec, "Integrated m-sequence based transceiver for uwb sensor networks," *Radioengineering*, vol. 28, no. 1, pp. 175–182, 2019.

[4] M. Sokol, P. Galajda, and P. Jurik, "A high-speed differential frequency divider-by-7 with a 50% output duty cycle," in *2022 32nd International Conference Radioelektronika (RADIOELEKTRONIKA)*. IEEE, 2022, pp. 1–5.

[5] J. Rovňáková and D. Kocur, "Short range tracking of moving persons by uwb sensor network," in *2011 8th European Radar Conference*. IEEE, 2011, pp. 321–324.

[6] M. Ricci, L. Crocco, and F. Vipiana, "A microwave imaging device for detecting contaminants in water-based food products," in *2022 16th European Conference on Antennas and Propagation (EuCAP)*. IEEE, 2022, pp. 1–3.

[7] F. Khan, A. Ghaffar, N. Khan, and S. H. Cho, "An overview of signal processing techniques for remote health monitoring using impulse radio uwb transceiver," *Sensors*, vol. 20, no. 9, p. 2479, 2020.

[8] "Dielectric resonators." [Online]. Available: http://www.t-ceram.com/dielectric-resonators.htm

[9] "Datasheet bfr740l3rh - infineon datasheet," Sep 2018. [Online]. Available: https://www.infineon.com/dgdl/Infineon-BFR740L3RH-DS-v03_00-EN.pdf?fileId=5546d46265f064ff01663896ffd44ecb

[10] "Datasheet 20 / 24 ghz super low noise ce3520k3 cel."

[11] A. Farr, G. Blackie, and D. Williams, "Novel techniques for electronic tuning of dielectric resonators," in *1983 13th European Microwave Conference*. IEEE, 1983, pp. 791–796.

[12] P. Jurík, "Diplomová práca: Návrh oscilátorov pre uwb senzorove systémy," 2021.

[13] "phase detector/frequency synthesizer data sheet adf4002 - analog devices." [Online]. Available: https://www.analog.com/media/en/technical-documentation/data-sheets/ADF4002.pdf

Overview of trends in the field of detection of hate speech and offensive language on social media

[1]*Zuzana SOKOLOVÁ (2nd year),*
Supervisor: [2]*Jozef JUHÁR*

[1,2]Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

[1]zuzana.sokolova@tuke.sk, [2]jozef.juhar@tuke.sk

***Abstract*—In this paper, we describe recent trends in the detection of hate speech and offensive language on social media. Based on the latest studies and scientific contributions, we describe the most used methods and achieved results in connection with the detection of hate speech and offensive language. We point out the topicality of the selected topic, describe the next direction of our work, and suggest possible solutions to current problems in this field of research.**

***Keywords*—hate speech detection, natural language processing, offensive language, social networks**

I. Introduction

Nowadays, social media has become a hotbed of hate speech that can have harmful effects on society. Hate speech or offensive language can damage relationships between different groups of people. Detecting hate speech and offensive language is important because it can help prevent these harmful effects. Negative thinking can lead, for example, to suicide, self-harm, psychological disorders, imbalance, stress-related illnesses, drug or alcohol addiction, and others [1]. Therefore, in our opinion, it is important to deal with this issue and at least partially limit the hateful and offensive language expressions of such people. It is for this reason that we want to focus on the detection of hate speech and offensive language on the Internet and social networks as part of the dissertation work.

The world of social media is currently saturated with unfiltered content, from cyberbullying and cyberstalking, to hate speech and offensive language. For this reason, identifying and cleaning up such toxic language is a major challenge and is currently a very active area of research. The arrival of Web 2.0[1] had a great impact not only on the availability of information but also on its generation. Therefore, addressing any aspect of hate speech on social media platforms involves applying deterrent and preventive mechanisms. Community or social networking sites have a warning in the standards that they prohibit toxic behavior on their platforms and encourage users to report such actions to platform administrators. While this measure (i.e. blocking users who have been reported for hate speech) is effective in addressing such acts, victimization cannot be prevented. Therefore, as part of prevention, all users must participate in filtering hate speech. At the same time, user-reported hate speech content should be checked to see if it is truly toxic to ensure users' freedom of expression is respected. However, the process of filtering out inappropriate reported content is done manually by relying on human editors. This ineffective technique takes time and certainly cannot keep up with the increasing toxic content on social media. However, platforms such as Facebook and YouTube still use a primitive method to filter hate speech based on content filtering using a specific dictionary that contains words most often associated with hate speech [3]. For this reason, the automatic detection of hate speech has become an urgent need, since the human factor cannot cope with such a large amount of emerging texts [4].

[1]Web 2.0 describes the current state of the internet, which has more user-generated content and usability for end-users compared to its earlier incarnation, Web 1.0. [2]

II. The current state of the research

The lack of suitable mechanisms for solving the problem of hate speech in the online environment has prompted several researchers to work on tasks aimed at automatically detecting hateful content. In several recently published papers, various approaches have been proposed for designing automatic hate speech text detectors. Existing studies for the detection of inappropriate textual content use data-driven approaches and supervised learning methods. These methods include high-performance machine learning (ML) and deep learning (DL) algorithms to create models that can detect the presence of hate speech. ML algorithms driven by different combinations of features characteristic of comments or users were used to build such models. DL algorithms have recently been used in the work on solving toxicity, while the authors have proposed different DL models built on convolutional neural networks and recurrent neural networks. The NLP community came up with other new techniques. One of them is the attention mechanism [5] or large-scale transfer learning [6] for solving hate speech detection tasks.

Many of the recent studies in this area have confirmed that detecting hate speech is a very difficult and complicated task. The authors in these posts indicated that there are several problems to be faced when dealing with textual content of this type. One of these problems is non-vocabulary words, which occur mainly in texts posted on social networks. Social media or media users often use grammatically incorrect words in their posts and comments. Also, many users misspell words on purpose to mask inappropriate words. Precisely to avoid algorithms designed to filter hate speech, which is based on the lexicon. Another problem is the imbalanced classes in the available hate speech datasets. Precisely because these data are relatively few in the text and are often not publicly available.

Avoiding bias when verifying the presence of hate speech in a text is also a problem. Because if we want to effectively and efficiently recognize the presence of hate speech in the text, it is necessary to improve the capabilities of the detectors so that they can understand the meaning of words in different contexts [3].

In the theoretical part of the dissertation work, we characterized in more detail all available methods designed to detect hate speech and offensive language in the text. Based on the studied issue, we found that the best results were achieved by word embedding methods in combination with deep learning, a separate method of long short-term memory [7], but also in combination with other methods [8]. Remarkable results were also achieved using association rule [9], Fuzzy logic [10], and a combination of the decision tree, genetic algorithm, and particle swarm optimization [11]. Conversely, worse results were achieved using the support vector method [12] and Bayesian networks [13]. In the studies, the authors most often devoted themselves to two tasks the detection of hate speech and offensive language. First, they dealt with the initial classification of the datasets they chose or created. They sorted the dataset into two or three classification classes. They divided the collected tweets, messages, or comments into those that contained hate speech/offensive language and those that did not. If they chose to classify the dataset into three classes, the most common categories were: hateful, offensive, or neutral. In the mentioned studies, the authors most often dealt with the problems of hate speech and offensive language in connection with race, religion, homophobia, and immigrants. From our point of view, the research and studies mentioned have been less concerned with the detection of hate speech and offensive language related to sexism, feminism, misogyny, gender equality, and toxic masculinity. As these categories have received very little research coverage, we would like to focus on them in our further research.

Most language models are pre-trained for general language regardless of the domain and style of language expression, as they are trained on the largest possible corpus of text data obtained from web sources. Based on previous scientific research, we believe that it would be appropriate to examine individual models, also in cooperation with the Question Answering technique, to see if they are too biased toward certain types of hate speech, such as racial or religious intolerance, sexism, feminism, misogyny, etc., and could thus contain elements of stereotypical expressions produced by the content creators themselves, either in the form of blogs on various discussion web forums or in the form of posts and comments on social networks. We would also like to focus on the creation of a corpus of text data for the detection of hate speech and offensive language in the Slovak language.

From our point of view, an interesting possibility would be to create a corpus similar to the StereoSet dataset for the Slovak language, which is described in detail in studies [14]. M. Nadeem et al. developed the Context Association Test (CAT) to measure stereotype biases of pre-prepared language models concerning their language modeling ability. They created a dataset called StereoSet that contained 16,995 CATs to test for bias in four categories: gender, race, religion, and professions. They showed that the current pre-trained language model exhibits strong stereotypic biases and that the best model is 27.0 ICAT points behind the idealistic language model.

III. Conclusion

Based on the knowledge of the current state of the automatic detection of hate speech and offensive language in social networks, we want to focus our dissertation work on several tasks. One of the tasks will be to analyze available language models based on transformers and verify their applicability in the automatic detection of hate speech and offensive language in the Slovak language. The second task is to design and create our database of hate speech and offensive language for the Slovak language (benchmark dataset), especially to evaluate language models from the point of view of their stereotypical bias using standard evaluation measures. The third task is to design and implement a new or modify existing approach to the automatic detection of hate speech and offensive language using transfer learning, based on the training of language models for the task of text classification.

Acknowledgment

The research in this paper was partially supported by the Ministry of Education, Science, Research and Sport of the Slovak Republic under the research project VEGA 2/0165/21, by the Slovak Research and Development Agency under the project of bilateral cooperation APVV SK-TW-21-0002, and by the Faculty of Electrical Engineering and Informatics, TU Košice under the Grant FEI-2023-95.

References

[1] Whitlock, Janis, Peter A. Wyman, and Sarah R. Moore (2014) Connectedness and suicide prevention in adolescents: Pathways and implications. *Suicide and Life-Threatening Behavior*, Vol. 44, No. 3, ISSN 1943-278X, doi:10.1111/sltb.12071

[2] Kenton W. (2022). What Is Web 2.0? Definition, Impact, and Examples. Available Online: https://www.investopedia.com/terms/w/web-20.asp

[3] Mazari, Ahmed Cherif, Nesrine Boudoukhani, Abdelhamid Djeffal (2023) BERT-based ensemble learning for multi-aspect hate speech detection. Cluster Computing, 1-15.

[4] Gao, L., Huang, R. (2017). Detecting online hate speech using context aware models. arXiv preprint arXiv:1710.07395.

[5] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... and Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30.

[6] Toraman, C., Şahinuç, F., and Yılmaz, E. H. (2022). Large-scale hate speech detection with cross-domain transfer. arXiv preprint arXiv:2203.01111.

[7] Wang, B. & Ding, H. (2019, June). YNU NLP at SemEval-2019 task 5: Attention and capsule ensemble for identifying hate speech. In Proceedings of the 13th International Workshop on Semantic Evaluation (pp. 529-534).

[8] Alarifi, A., Tolba, A., Al-Makhadmeh, Z., & Said, W. (2020). A big data approach to sentiment analysis using greedy feature selection with cat swarm optimization-based long short-term memory neural networks. The Journal of Supercomputing, 76, 4414-4429.

[9] Wadhwa, P., & Bhatia, M. P. S. (2014). Classification of radical messages in Twitter using security associations. Case studies in secure computing: Achievements and trends, 273-294.

[10] Liu, H., Burnap, P., Alorainy, W., & Williams, M. L. (2019, May). Fuzzy multi-task learning for hate speech type identification. In The world wide web conference (pp. 3006-3012).

[11] Nagarajan, S. M., & Gandhi, U. D. (2019). Classifying streaming of Twitter data based on sentiment analysis using hybridization. Neural Computing and Applications, 31, 1425-1433.

[12] Perelló, C., Tomás, D., Garcia-Garcia, A., Garcia-Rodriguez, J., & Camacho-Collados, J. (2019, June). UA at SemEval-2019 task 5: setting a strong linear baseline for hate speech detection. In Proceedings of the 13th International Workshop on Semantic Evaluation (pp. 508-513).

[13] Chakravartula, N. (2019, June). Hatemiter at SemEval-2019 task 5: hate speech detection against immigrants and women in Twitter using a multinomial naive Bayes classifier. In Proceedings of the 13th International Workshop on Semantic Evaluation (pp. 404-408).

[14] Nadeem, M., Bethke, A., and Reddy, S. (2020). StereoSet: Measuring stereotypical bias in pretrained language models. arXiv preprint arXiv:2004.09456.

Dynamic allocation of Multi-Access communication and computing resources of mobile users at the edge using Artificial Neural Networks

[1]*Matúš ČAVOJSKÝ (1st year),*
Supervisor: [2]*Gabriel Bugár*

[1,2]Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

[1]matus.cavojsky@tuke.sk, [2]gabriel.bugar@tuke.sk

***Abstract*—Mobile devices, autonomous vehicles and IoT increasingly rely on artificial neural networks (ANNs) to execute complex inference tasks. Constant execution of the entire neural network on a mobile device can quickly deplete its battery or become a burden for the integrated processor. Although task offloading to edge servers may decrease the mobile device's computational load, unpredictable channel quality, network and edge server load may lead to a significant delay in task execution. Recent approaches based on split computing (SC) have been proposed where the ANN is split into two models. A so-called head model is running on the mobile device and a tail model, executing on the edge server if available, possibly reducing bandwidth usage as well as energy consumption. Existing research also suggests presenting multiple "exits" earlier in the architecture, called early exiting (EE), each providing increasingly higher target accuracy. A trade-off between model accuracy and execution time can be tuned according to the current requirements or application demands.**

***Keywords*—Artificial Neural Networks, Early Exit, Multi-access Edge Computing, resource allocation, Split Computing**

I. Introduction

Rapid evolution in computing systems has continuously reduced their form factor from a room-size mainframe to a miniature battery-powered Internet-of-Things (IoT) device with centimeter/millimeter-scale size [1]. The miniature system becomes an attractive monitoring solution by continuously sending, receiving, and analyzing data in a feedback loop. Analysis can be conducted either by humans or artificial neural network (ANN), in near real-time or over a longer period. With the proliferation of 5G mobile networks, various immersive applications such as self-driving cars, smart homes, smartwatches, virtual and augmented reality, and the industrial IoT are penetrating into our daily routines. Within the IoT category, Cisco estimated that the connected home applications will have the largest share and connected car will be the fastest growing application type [2]. Since most of these services require the transmission of massive volumes of data and extremely complex data processing in the cloud, the existing network and cloud infrastructures are likely to be placed under high pressure in the foreseeable future [3]. In recent reports, Cisco has estimated approximately 20 zettabytes of yearly data center traffic by 2021 [4].

II. Related works

The high-rate and highly-reliable air interface allows to run computing services of mobile devices at the remote cloud data center, resulting in the research area called mobile cloud computing [5]. One of the most important challenges in mobile cloud computing is to address the contradiction between the increasing complexity of mobile applications and the limited local capabilities of mobile devices. A viable solution to this challenge is to leverage cloud computing and execute mobile applications remotely [6].

A traditional cloud computing paradigm, in which all requests are processed at a number of large data centers placed at multiple geographic locations around the world. Thus, requests and responses are forwarded to a remote server through multiple segments of the transport network infrastructure. With the rapid increase in data volumes due to rich multimedia applications and the deployment of massive numbers of IoT devices, the transmission of context-aware traffic from multiple areas to the same remote cloud servers is expected to occupy a sufficient share of the total network traffic to negatively impact network efficiency [3].

Therefore, there are pressing needs to redesign the cloud architecture to serve the mobile workloads with better efficiency and scalability. Recent research efforts propose to deploy an intermediate cloud layer, so-called cloudlets, at the network edge [6].

A. Edge computing

Satyanarayan et al. [7], [8] introduced the concept of cloudlet as the middle layer in a three-tier architecture consisting of an end-device, an edge cloud platform, i.e., cloudlet, and a centralized datacenter. The objective of cloudlet is to extend the remote datacenter cloud services in close proximity to the end users. Cloudlet is viewed as a trusted and resource-rich node with stable Internet connectivity, offering computing as well as access and storage resources to nearby mobile devices. It was proposed as an edge cloud node, which can reside in community places [9]. The researchers chose the first layer of the cloudlet design as a new way of offloading computationally demanding and time-critical tasks, which was named edge computing (EC). As shown in Figure 1, edge computing focuses on task computation on the first cloudlet note - the closest to user. Since the distance between mobile

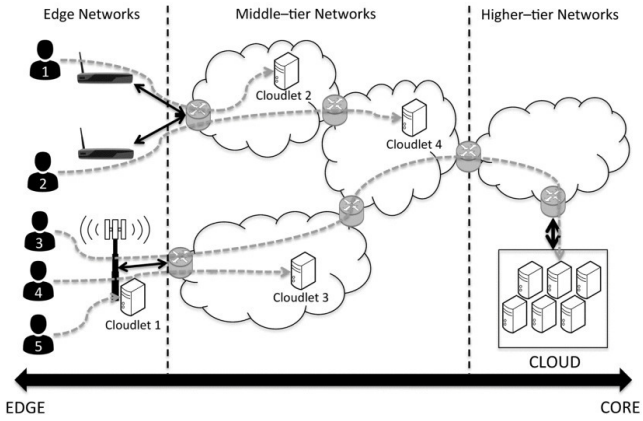


Fig. 1. Multi-tier edge computing diagram of users requesting service by cloudlets at the edge and middle-tier locations of the network [10].

devices and the cloud is shortened, the expensive cloud-to-edge data transmission costs and the network latency accessing cloud computing services could be significantly reduced.

Autonomous vehicles are an example of how can edge computing and IoT work together. An autonomous vehicle needs to collect and process real-time data about traffic, pedestrians, street signs and stop lights, as well as monitor the vehicle's systems. Sending, processing and receiving data from the remote cloud would take too long, if the vehicle needed to stop or turn quickly to avoid an unexpected obstacle. EC brings cloud computing services geographically closer to the vehicle, allowing the sensors in the vehicle to offload tasks to avoid an accident.

The European Telecommunications Standards Institute provided a concept of multi-access edge computing (MEC) [11]. In the MEC architecture, distributed edge servers are located at the network edge to provide computing capabilities and IT services with high bandwidth and real-time processing. Edge servers become the third offloading location of compute-intensive tasks in addition to end devices and cloud. However, due to edge servers' restricted computing capability, they cannot completely take place of cloud servers [12]. The performance of the services deployed on edge servers is dependant on edge network density, measured by the average number of links per edge server, allowing edge servers' to collaborate. In general, a high network density connects edge servers to many other edge servers and allows them to utilize more system resources through collaboration. This allows for high service performance because it enables low-latency messaging and data transmissions between edge servers [13].

However, canonical EC does not consider that the quality and availability of wireless links can suddenly fluctuate due to the presence of constantly changing noise levels and interference patterns, which may impair performance in latency-bound applications. For instance, TCP protocol interactions triggered by mobility and impaired propagation have been proven to induce erratic capacity patterns even in high-capacity 5G links [14]. Furthermore, most IoT devices do not support high data rates and instead rely solely on lower-power technologies such as LoRa or Wi-Fi. The severe offloading limitations of some mobile devices, coupled with the instability of the wireless channel and limited resource allocation of edge servers, imply that the amount of data offloaded to edge should be decreased as much as possible, while at the same time keep the result accuracy as close as possible to the original. For this reason, split computing (SC) [15] and early exiting (EE) strategies [16] have been proposed to provide an intermediate option between EC and local computing [17].

B. Split computing

One of the most common compute-intensive tasks is deep neural network (DNN) inference. A DNN is an ANN with multiple hidden layers between the input and output layers. As DNN-based classifiers improve their predictive accuracy, mobile applications such as speech recognition in smartphones, real-time unmanned navigation and drone-based surveillance are increasingly using DNNs to perform complex inference tasks [17]. On this type of task, many researchers study specialized computation offloading strategies. Kang et al. [15] proposes Neurosurgeon for DNN offloading. Neurosurgeon divides DNN into two parts, essentially splitting the model in half, while the architecture and weights of the head and tail model are exactly the same as in the original model. One part runs at end devices locally and the other runs at the cloud. This simple approach preserves accuracy but allocates part of the execution to the mobile device, whose computing power is expected to be smaller than that of the edge server, so that the total execution time may be larger while preserving battery life. The neurosurgeon method reduces the calculation at end devices, trying to find a balanced point between computation and transmission time [12]. Inference time in SC is the sum of processing time on mobile device, delay of communication between mobile device and edge server (uplink of the task), the processing time on edge server and delay of response between edge server and mobile device (receiving of the solution).

Inference on edge can be achieved either by running computations on CPU or on hardware accelerators, such as dedicated GPU, GPGPU, DSP or using dedicated neural processing engines. The benefits and demand for on-device machine learning is driving modern phones to have dedicated neural engine or tensor processing units. For example, Apple iOS 15 will support on-device speech recognition for iPhones with Apple neural engine [18].

Splitting model accomplishes a task-oriented compression guarantying a higher degree of privacy compared to sending sensor data directly to edge server. As a matter of fact, no encryption is necessary as the representation (output from neural network) may lack information needed to fully reconstruct the original input data, further reducing the load on end device. The main objective of SC is achieving distribution of computing load among a mobile device and edge server and establishes a task-oriented compression to reduce data transfer delays.

In order to reduce load on edge servers a novel approach to modeling a deep neural network can interrupt inference if the confidence level of neural network is high enough that additional computation would be almost meaningless.

C. Early Exiting

Another approach to enable mobile computing is referred to early exiting. A conventional neural network inference is performed with the same operation regardless of whether the image is easy to classify or difficult. The core idea of EE, first proposed in Teerapittayanon et al. [16], is to circumvent the need to make DNN models smaller by introducing early exits in the DNN, where execution is terminated at the first

exit achieving the desired confidence on the input sample. For instance, some samples in test datasets (and in real-world problems) will be easy for a DNN model, but others may not be, depending on ML models we use. Thus, EE ends the inference process with fewer transforms (layers) for such easy samples so that the overall inference time and computation cost are reduced [17]. This means that the added capacity is redundant for those samples. Each inference data (e.g., an image) is processed sequentially by the N-1 early exits until an inference result with sufficient confidence is obtained. It is also possible that none of the early exits is able to obtain a reliable inference result, and in this case, the accuracy is determined by the last (Nth) exit [19].

Teerapittayanon et al. [20] also applies this idea to mobile-edge-cloud computing systems, where the smallest neural model is allocated to the mobile device. If a confidence threshold for the input is not reached, the intermediate output is forwarded to the edge server, where inference will continue using a mid-sized neural model with another exit. Lastly, intermediate layer's output is forwarded to the cloud, which executes the largest neural model, if the output still does not reach the target confidence.

Figure 2 provides an overview of local, edge, split computing and early-exit models, which are the main computational paradigms discussed in the paper.

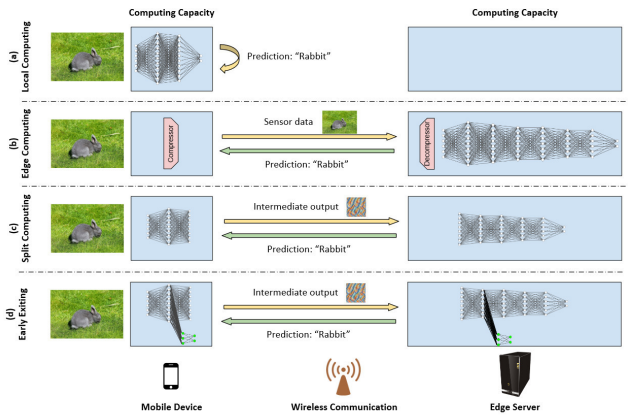


Fig. 2. Overview of (a) local, (b) edge, (c) split computing, and (d) early exiting: image classification as an example [17].

To achieve more efficient inference without significantly sacrificing the accuracy of the original model, the system needs to find balance between classifiers. As recent studies introduce multiple early exits to a model at different stages, such optimizations is challenging [17]. The higher the confidence values of the early classifiers is set, the more samples will not be stopped at the end device. The system needs to find balance between overall accuracy and amount of data sent to the edge node.

D. Advantages and disadvantages

The advantages of executing compute-intensive tasks on end devices are twofold. On the one hand, sensor data, e.g., image, audio and video, are generated at end devices. Processing data locally on end devices can avoid time-consuming data transmission and reduce heavy bandwidth consumption, compared to sending these data to the cloud server. On the other hand, some tasks are sensitive to latency and the execution result can be out of date if being late. In some cases, e.g., face recognition applications, high latency can result in poor user experience. The unreliable and delay-significant wide-area connection can be problematic if computation tasks are to be offloaded to the cloud or the network edge. Therefore, executing compute-intensive tasks on end devices is a potential solution to lower end-to-end latency.

E. Performance analysis of EE model

The first model using early exists was BranchyNet [16] used for image classification of CIFAR-10 and MNIST datasets. Early exists were used to improve inference time and reduce the number of samples for every stage of the neural network. LeNet evaluated on MNIST dataset achieved early exit sample percentage of 94.3% and only 5.63% samples needed to be processed by whole network. AlexNet implementation evaluated on CIFAR-10 dataset had two early exits accomplishing first early exit sample percentage of 65.6%, second EE of 25.2% and only 9.2% used whole networks' layers. Exiting these samples early translate to CPU/GPU speedup gains of 5.4/4.7x over LeNet, 1.5/2.4x over AlexNet [16]. The need to send only a small percentage of samples to edge node could improve it's availability and ultimately reduce the power cost of sending and receiving data from the network.

III. Research Challenges

Existing studies of MEC infrastructure focus on the placement of edge servers across optional locations to achieve various optimization objectives, e.g., minimum edge server deployment cost [21], cost-effective resource allocation [3], spot pricing [10], maximum user coverage, maximum edge network robustness, etc. However, the importance of EC node availability in split computing scenarios and scenarios with EC node outage is neglected. None of the models in the survey [17] used split models of motion detection and/or prediction. Autonomous vehicles are an example why motion detection is so important to path planning and decision making.

Motion detection is important in path planning and decision-making for autonomous vehicles because an autonomous vehicle needs to collect and process real-time motion data, preferably from cameras, and find a solution (the path) in a timely fashion. By detecting and analyzing the motion of objects in its surroundings, the vehicle can make informed decisions about its trajectory and avoid potential collisions. Motion detection helps the vehicle to understand the movements of other objects, such as pedestrians and other vehicles, and to plan its own movements accordingly. This is critical for ensuring the safety of passengers and other road users. Existing Mobile Edge Computing (MEC) infrastructure placement models address the importance of motion detection in distributed computing scenarios by considering the mobility patterns and traffic demand of mobile devices. For instance, some models take into account the spatial and temporal correlations of the motion data generated by the devices, and use this information to optimize the placement of MEC servers [22].

Unsolved problems in SC and EE include:

1) Optimization of computation distribution: While Split Computing offers advantages in terms of reducing the workload on central servers, it can also introduce complexities in terms of how computations are distributed

across multiple devices. Optimization of this process remains an active area of research.
2) Resource allocation: Allocating resources among different computing devices is a challenging problem in Split Computing. Determining the optimal allocation of resources to each device requires the consideration of several factors, such as device capabilities, network bandwidth, and the type of computation being performed.
3) Security: Split Computing introduces new security concerns, particularly in terms of data privacy and confidentiality. Ensuring that data has not been tampered with when being transmitted between devices is a key area of research.
4) Fault tolerance: When multiple devices are involved in a computation, there is an increased risk of device failure. Developing strategies for detecting and responding to device failures is an ongoing area of research.
5) Incorporating multiple exit points: Most current early exit models use a single exit point, but incorporating multiple exit points is an area of active research. Determining the optimal number and placement of exit points can be challenging and may depend on the specific application.
6) Optimization of the early exit criteria: Determining the optimal criteria for making an early exit decision is an active area of research. The criteria used to determine whether a prediction should be made early or not can have a significant impact on the overall performance of the model.
7) Generalization to different datasets: Early exit models can be particularly sensitive to imbalanced data, where there are unequal numbers of examples in different classes. Changes in the distribution of the data, can make it challenging for model to generalize to new datasets. Developing strategies for handling imbalanced data and improving the generalization in early exit models is an ongoing area of research.

Solvability of these problems is dependant on their nature, some of them are likely solvable with further research and development. For example, optimization of computation distribution, resource allocation and optimization of the early exit criteria are problems that are likely solvable with continued research and development, while models incorporating multiple exit points already exist. On the other hand, security, fault tolerance and generalization to different datasets are problems that may be more challenging to solve completely, as they are affected by a wide range of factors that are difficult to control completely.

Existing research from Matsubara et al. suggests further research to focus on:

1) Evaluation of SC and EE in more practical settings: Create a SC ANN model an evaluate the SC and EE performance on a real-world device, not on a dataset.
2) Optimization of bottleneck design and placement in SC: Evaluate bottleneck data size (or compression rate), complexity of head model executed on mobile device, and resulting model accuracy (as higher compression rate may impact accuracy significantly).
3) Dynamic control of exits in EE: Although Ju et al. [23] proposed dynamic early exit scheduling through contextual bandits, more research should be done is this area.
4) Expanding the Application Domain of SC and EE to include more types of ANNs, i.e. natural language processing or health monitoring as these tasks are much more sensitive to prediction accuracy and average inference time.

Although extensive research studies have been done in the area of placement of EE, novel approach to early exit modeling is to place EE as close to end of the head model as possible and save results from EE even if they are of lower accuracy than large model could deliver. Results can be saved and used if edge server is unresponsive, task timeout occurs or is communication interrupted. In case of server response the response is used and the difference in prediction output can be logged to be used for model improvement by training. PhD dissertation work will be focused on researching application of a SC ANN model in area of Two Generals' Problem - networking.

IV. Conclusion

IoT devices such as smartphones and drones have now become an integral part of our daily lives. These devices increasingly utilize artificial neural networks to execute complex inference tasks such as image classification and speech recognition, among others. Split computing and early exiting are new paradigms gaining attention from many researches focused on specialized computation offloading strategies. The paper also provides a set of compelling research challenges that could improve existing work in the field.

Acknowledgment

This work was supported by the Ministry of Education, Science, Research and Sport of the Slovak Republic, and the Slovak Academy of Sciences under Grant VEGA 1/0685/23; by the Slovak Research and Development Agency under Grant APVV SK-CZ-RD-21-0028 and from the Operational program Integrated infrastructure project ITMS 313011V422, co-financed by the European Regional Development Found.

References

[1] Y. Lee, G. Kim, S. Bang, Y. Kim, I. Lee, P. Dutta, D. Sylvester, and D. Blaauw, "23.2 a modular 1mm 3 die-stacked sensing platform with optical communication and multi-modal energy harvesting," *Digest of Technical Papers - IEEE International Solid-State Circuits Conference*, vol. 55, pp. 402–404, 11 2012.
[2] C. and/or its affiliates, "Cisco annual internet report (2018–2023) white paper," Cisco, Tech. Rep., 2020. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html
[3] E. Šlapak, J. Gazda, W. Guo, T. Maksymyuk, and M. Dohler, "Cost-effective resource allocation for multitier mobile edge computing in 5g mobile networks," *IEEE Access*, vol. 9, pp. 28 658–28 672, 2021.
[4] C. and/or its affiliates, "Cisco global cloud index: Forecast and methodology, 2016–2021 white paper," Cisco, Tech. Rep., 2018. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/global-cloud-index-gci/white-paper-c11-738085.html
[5] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "A survey on mobile edge computing: The communication perspective," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 4, pp. 2322–2358, 2017.
[6] L. Tong, Y. Li, and W. Gao, "A hierarchical edge cloud architecture for mobile computing," in *IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications*, 2016, pp. 1–9.
[7] M. Satyanarayanan, P. Bahl, R. Caceres, and N. Davies, "The case for vm-based cloudlets in mobile computing," *IEEE Pervasive Computing*, vol. 8, no. 4, pp. 14–23, 2009.

[8] M. Satyanarayanan, "Fundamental challenges in mobile computing," in *Proceedings of the Fifteenth Annual ACM Symposium on Principles of Distributed Computing*, ser. PODC '96. New York, NY, USA: Association for Computing Machinery, 1996, p. 1–7. [Online]. Available: https://doi.org/10.1145/248052.248053

[9] T. Taleb, K. Samdanis, B. Mada, H. Flinck, S. Dutta, and D. Sabella, "On multi-access edge computing: A survey of the emerging 5g network edge cloud architecture and orchestration," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 3, pp. 1657–1681, 2017.

[10] A. G. Tasiopoulos, O. Ascigil, I. Psaras, S. Toumpis, and G. Pavlou, "Fogspot: Spot pricing for application provisioning in edge/fog computing," *IEEE Transactions on Services Computing*, vol. 14, no. 6, pp. 1781–1795, 2021.

[11] E. T. S. Institute. (2023) Multi-access edge computing - standards for mec. [Online]. Available: https://www.etsi.org/technologies-clusters/technologies/multi-access-edge-computing

[12] C. Haowei, Z. Liekang, Y. Shuai, and C. Xu, "Knowledge distillation for mobile edge computation offloading," *CoRR*, vol. abs/2004.04366, 2020. [Online]. Available: https://arxiv.org/abs/2004.04366

[13] R. Luo, H. Jin, Q. He, S. Wu, and X. Xia, "Cost-effective edge server network design in mobile edge computing environment," *IEEE Transactions on Sustainable Computing*, vol. 7, no. 4, pp. 839–850, 2022.

[14] M. Zhang, M. Polese, M. Mezzavilla, J. Zhu, S. Rangan, S. Panwar, and M. Zorzi, "Will tcp work in mmwave 5g cellular networks?" *IEEE Communications Magazine*, vol. 57, no. 1, pp. 65–71, 2019.

[15] Y. Kang, J. Hauswald, C. Gao, A. Rovinski, T. Mudge, J. Mars, and L. Tang, "Neurosurgeon: Collaborative intelligence between the cloud and mobile edge," *SIGARCH Comput. Archit. News*, vol. 45, no. 1, p. 615–629, apr 2017. [Online]. Available: https://doi.org/10.1145/3093337.3037698

[16] T. Surat, M. Bradley, and K. H. T., "Branchynet: Fast inference via early exiting from deep neural networks," *CoRR*, vol. abs/1709.01686, 2017. [Online]. Available: http://arxiv.org/abs/1709.01686

[17] Y. Matsubara, M. Levorato, and F. Restuccia, "Split computing and early exiting for deep learning applications: Survey and research challenges," 2021. [Online]. Available: https://arxiv.org/abs/2103.04505

[18] S. Gondi and V. Pratap, "Performance evaluation of offline speech recognition on edge devices," *Electronics*, vol. 10, no. 21, 2021. [Online]. Available: https://www.mdpi.com/2079-9292/10/21/2697

[19] R. Dong, Y. Mao, and J. Zhang, "Resource-constrained edge ai with early exit prediction," *Journal of Communications and Information Networks*, vol. 7, no. 2, pp. 122–134, 2022.

[20] S. Teerapittayanon, B. McDanel, and H. Kung, "Distributed deep neural networks over the cloud, the edge and end devices," in *2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS)*, 2017, pp. 328–339.

[21] F. Zeng, Y. Ren, X. Deng, and W. Li, "Cost-effective edge server placement in wireless metropolitan area networks," *Sensors*, vol. 19, no. 1, 2019. [Online]. Available: https://www.mdpi.com/1424-8220/19/1/32

[22] L. Ale, N. Zhang, S. King, and J. Guardiola, "Spatio-temporal bayesian learning for mobile edge computing resource planning in smart cities," *ACM Transactions on Internet Technology*, vol. 21, 03 2021.

[23] W. Ju, W. Bao, L. Ge, and D. Yuan, "Dynamic early exit scheduling for deep neural network inference through contextual bandits," in *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, ser. CIKM '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 823–832. [Online]. Available: https://doi.org/10.1145/3459637.3482335

Computer Vision Services in Transportation

[1]*Kristián MIČKO (2nd year),*
Supervisor: [2]*Peter PAPCUN*

[1,2]Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1]kristian.micko@tuke.sk, [2]peter.papcun@tuke.sk

***Abstract*—Camera systems and connected computers can collect and process data suitable for the intelligent transportation system. Motion detection and object tracking algorithms implemented in the camera system can count and categorize the vehicles on the highway. Our research focuses on using traditional computer vision methods, enabling us to use their computation on edge computing in devices such as microcontrollers or single-board computers. We focused on analyzing and implementing service-oriented architectures for reliable computation distribution of computer vision methods.**

***Keywords*—traditional computer vision methods, intelligent transportation, service-oriented architecture, edge computing.**

I. Introduction

We use motion detection and object tracking to analyze the activity on the road. The observation process of motion detection and object tracking is possible with cameras or intrusive sensors such as inductive loops or magnetometers. The intrusive sensors are very accurate for vehicle motion detection and vehicle categorization. These sensors installation is costly and also increases the maintenance cost. The camera system also can detect the vehicle's motion and categorize them as well as observe the vector of the vehicle's movement (Object Tracking) and automatically processes the license plates with ALPR (automatic license plate recognition) system.

This paper will summarise the suitable architectures for computer vision methods in intelligent transportation systems. After analyzing suitable architectures of computer vision methods deployment, we will analyze the suitable camera surveillance systems connected to edge devices such as microcontrollers (MCU) or single-board computers (SBC).

II. The initial status

In the previous analysis [1], we focused on the methods used in computer vision for motion detection and object-tracking problem in an intelligent transportation system. To solve motion detection and object tracking tasks via camera systems, we can use traditional computer vision or deep learning-based approaches [2]. The research challenge in optimizing the suitable one depends on the computational architecture. The complexity of the computer vision methods impacts the computational power and implementation time [3]. The deep learning (DL) approach is more suitable for complex problems that lack deep domain knowledge with extensive dataset collection, as well as having much computational power for the training and deployment of the DL model. The traditional approach is suitable for using domain knowledge to extract key classification features [4].

We want to divide feature extraction steps in motion detection, object-tracking, and ALPR methods into independent services and plugins from the view of the software product. Independent services could be deployable across many computation nodes, enabling us to distribute the services between edge, fog, and cloud depending on their computation complexity. The optimal computation distribution between edge, fog, and cloud computing relies on the power of SBC/MCU or desktop/server computers as well as the communication protocols from the software and hardware layers. Our research focuses on choosing the optimal communication method between computation nodes for service deployment and finding the suitable power of SBC/MCU for using these services.

III. The tasks solved in previous year

Our research focuses on using traditional computer vision methods because it enables them to be deployable in various SBC/MCU. Motion detection and object-tracking tasks have appropriate complexity to solve with the traditional computer vision method. We tested edge detectors combined with template matching to solve ALPR tasks and vehicle detection and counting. The measured computational consumption of mentioned methods is suitable for SBC/MCU like STM32 Nucleo or Raspberry Pi.

A. The architectures overview

We can divide the camera surveillance system into static and dynamic. The static surveillance system prefers wired connection between cameras and computers. This type of connection is selected in static highway tolls because electric power and optical fiber are available on highway roads [5]. Outside of the highways, the police also use the mobile camera system for ALPR or object-tracking purposes via drones. The drone is an example of a dynamic camera surveillance system because the video transmission is wireless, and the camera in the drone can change its position [6]. We describe the advantages and disadvantages of dynamic or static camera systems.

We analyzed many architectures for the smart surveillance system in intelligent transportation systems via drones [7] or static highway tolls [8], and IoT techniques [9], [10].

The analyzed computational architectures in the previous year inspired us to propose, implement and test our architecture shown in Fig 1. We used a combination of many SBCs to preprocess image data for motion detection via a background subtractor. The background subtractor crops the image in the area of movement. The SBC with implemented

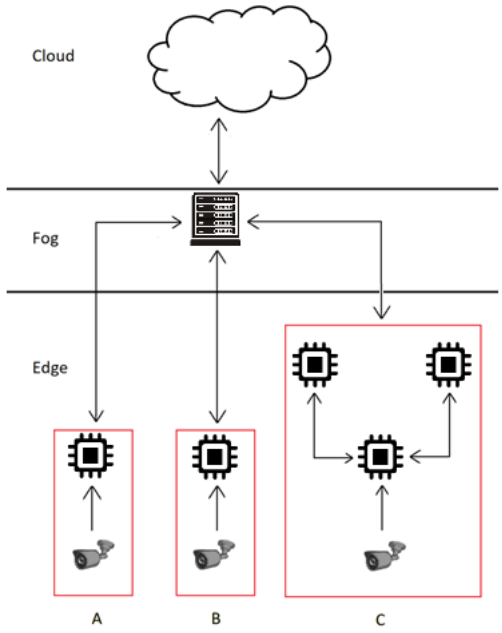


Fig. 1. The implemented architecture in our research

background subtractor sends the cropped image to another SBC via REST API protocol encoded with base64 format. The second SBC holds saved the template collection used for the vehicle classification via the template matching method. The template matching method calculates the probability of the occurrence of the template object in the cropped image. We use a license plate as a template collection because every vehicle has to carry them. If the template matching shows the probability of occurrence of a license plate of more than 80%, we classify the vehicle's presence in the cropped image. After the positive classification of the vehicle, we send a signal to the third SBC to run the MOSSE tracker to calculate the movement vector. The calculated and gained information are sent to the fog or cloud computing.

B. Services and plugins

The applications can be developed by monolithic architecture or service-oriented architecture. The monolithic architecture defines the creation of the application in one programming language, and the software based on this architecture does not use any computational distribution. The complex software created with the monolithic architecture requires powerful hardware. In the service-oriented architecture, the programmers compose the software from independent microservices that can be created from various programming languages, enabling the distribution of the microservices across many computational devices [11].

Each device can represent a computational node responsible for a particular algorithm or method. We focus on the computer vision methods deployed on the SBC or MCU. The SBC and MCU have limited computational resources, and complex computer vision programs require powerful hardware or some computational distribution across several low-powerful devices. Therefore we propose to research the service-oriented architecture deployment of computer vision methods and algorithms into SBC or MCU.

The service-oriented architecture we tried to applicate in the ALPR system on the highway toll. The motion detection algorithm runs on low-powerful SBC. This SBC takes a photo from a highway toll's video stream when it detects the vehicle's motion. The photo is processed and sent to the other device via HTTP protocol. That device has implemented an OCR neural network model that requires more computational time to read characters from the license plate.

We also implemented another plugin that counts vehicles in the selected road sections. The plugin counts vehicles with a motion detection algorithm based on the Canny edge detector and background subtractor. The SBC uses this plugin, where the SBC connected directly to the camera counts the vehicles. This SBC sends the count of the vehicles to the fog or cloud computing layer for the following analysis tasks. The object tracker microservice implemented in the other SBC also estimates a vector of movement of the observed vehicles and sends the direction information to the fog or cloud computers.

IV. Future work

This paper describes the service-oriented architectures that were used in our following research. Our research will focus on motion detection and object-tracking methods sensed from cameras on highway tolls via services and plugins.

We want to prove that the intelligent sensor system based on the combination of MCU/SBC, camera system and suitable computer vision methods can replace intrusive sensors for motion detection and object tracking tasks in the intelligent transportation system.

The testing of various MCU/SBCs shows which computer vision methods (traditional or DL based) are the optimal solution for making the intelligent sensor system. The intelligent sensor system will focus on motion detection and object tracking used for intelligent transportation system purposes.

Acknowledgment

This publication is the result of the APVV grant ENISaC - Edge-eNabled Intelligent Sensing and Computing (APVV-20-0247).

References

[1] M. Kristian, "Motion detection and object tracking in transportation based on edge computing," in *22nd Scientific Conference of Young Researchers : proceedings from conference.* TUKE, 2022.

[2] N. O'Mahony and et al., "Deep learning vs. traditional computer vision," in *Advances in Computer Vision*. Springer International Publishing, 2020.

[3] Y. Gong and et al., "Compressing deep convolutional networks using vector quantization," *CoRR*, 2014.

[4] "Decoding the dichotomy: Traditional image processing vs. deep learning." [Online]. Available: https://www.imveurope.com/sites/default/files/content/white-paper/pdfs/HCL_IMVE_WP-ImageProcessing_vs_DL.pdf

[5] A. Suryatali and V. Dharmadhikari, "Computer vision based vehicle detection for toll collection system using embedded linux," in *2015 International conference on circuits, power and computing technologies [ICCPCT-2015]*. IEEE, 2015.

[6] P. Teixidó and et al, "Secured perimeter with electromagnetic detection and tracking with drone embedded and static cameras," *Sensors*, 2021.

[7] N. Chen and et al., "Dynamic urban surveillance video stream processing using fog computing," 2016.

[8] Q. Zhang and et al., "Demo abstract: Evaps: Edge video analysis for public safety," in *2016 IEEE/ACM Symposium on Edge Computing (SEC)*, 2016.

[9] K. Micko and et al., "Temporary parking via computer vision and deep learning," in *2022 IEEE 20th (SAMI)*. IEEE, 2022.

[10] K. Micko and F. Babic, "Evidencia dočasného parkovania pomocou metód počítačového videnia a umelej inteligencie," *Electrical Engineering and Informatics XII*, 2021.

[11] J.-P. Gouigoux and D. Tamzalit, "From monolith to microservices: Lessons learned on an industrial migration to a web oriented architecture," in *2017 IEEE (ICSAW)*. IEEE, 2017.

Updated design of tool for automated optimization and parallelization in heliospheric field

[1]*Michal SOLANIK (3rd year),*
Supervisor: [2]*Ján GENČI, Consultant:* [3]*Pavol BOBÍK*

[1,2]Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic
[3]Department of Cosmic Physics, Institute of Experimental Physics SAS Kosice, Slovak Republic

[1]michal.solanik@tuke.sk, [2]jan.genci@tuke.sk, [3]bobik@saske.sk

***Abstract*—Models of cosmic rays distribution in the heliosphere are rarely published, with only two public implementations available - Geliosphere and SolarProp. Development of these models is very demanding on computing resources. To support the development of heliospheric models, we introduced the design of an optimization flow of tool for automated optimization and parallelization. We've found several flaws in the original design. This paper presents an updated design of the tool and verification method based on a prototype, which was able to produce efficient parallel GPU implementation.**

***Keywords*—Automated optimization, domain-specific languages, heliosphere, GPGPU**

I. Introduction

Multiple authors and collaborations develop models of cosmic rays distribution in the heliosphere. Despite of interest of the scientific community, only two public implementations of models are open-source and available to the public, SOLARPROP [1] and Geliosphere [2]. Development of models of cosmic rays distribution in the heliosphere is a slow process, especially verification of the implementation of the model can last days to get enough quantity of registered particles [3] to render a spectrum with 1% error on low energies. In [4] we presented the concept of a tool for automated optimization and parallelization in the heliospheric field. In this paper, we present a corrected version of the concept along with results from the prototype focused on basic optimization and parallelization in GPU implementation.

II. Original design of tool

Figure 1 presents the original design of optimization flow in a tool for automated optimization and parallelization in the heliospheric field. Flow is based on a genetic algorithm defined by Koza in [5]. For further clarification, in this design, we use recommendation as a word expressing changes, which should be applied to the domain model. Flow [4] can be divided into the following steps:

- Transformation of source code into domain model - input source code is transformed into the domain model, which allows easier manipulation in the further dynamic and static analysis,
- Static analysis - analysis generating recommendations based on the state of the domain model,
- Dynamic analysis - analysis generating recommendations based on output from external tools,
- Generate generation of programs based on recommendations - generator should take all recommendations and generate programs from the domain model with the combination of all non-conflicting recommendations,
- Evaluation of generation of programs - two forms were presented, evaluation based on execution time and evaluation based on external tools,
- Filter results - filter should have access to all evaluation data, to ensure the selection of the best fitting program or to create a new generation.

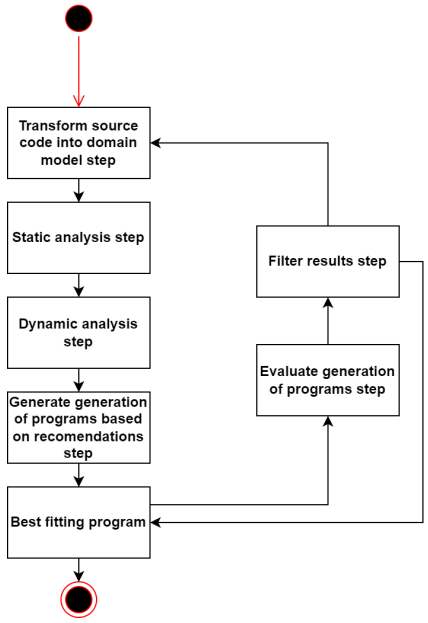


Fig. 1. Original design of optimization flow in tool for automated optimization and parallelization in heliospheric field [4]

During the further design of the tool, we encountered the following issues:

- Incorrect placement of dynamic analysis - dynamic analysis is based on output from external tools. Separation of dynamic analysis from the execution of external tools can complicate further extension of dynamic analysis,
- Inaccurate definition of input format and design - input format can be considered as an implementation detail, however, we should specify if we want to use domain-specific or general-purpose language,
- Incorrect availability to proceed from generating genera-

tion of programs into the best fitting program - responsibility to select the best fitting program should only belong to the filter step.

III. Updated design of tool

In section II we described several flaws contained in the original design of the tool. Figure 2 presents the updated design of optimization flow in a tool for automated optimization and parallelization in the heliospheric field, which reflects these flaws.

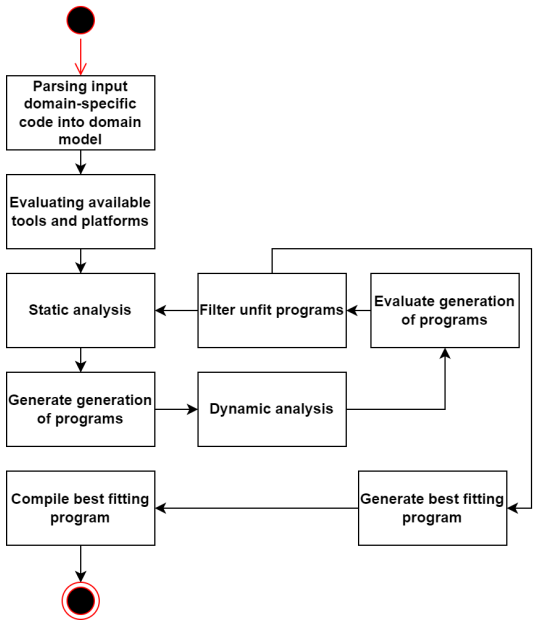


Fig. 2. Updated design of optimization flow in tool for automated optimization and parallelization in heliospheric field

The original design did not specify the definition of input. Usage of general-purpose languages can be difficult, especially in cases when we want to generate a program for another platform like GPU. The definition of domain-specific language is also more suitable for support of specific functionality needed in the heliospheric field, such as the injection of initial kinetic energy. This design change is reflected in the first step of the updated optimization flow presented in Figure 2.

The placement of the dynamic analysis step in the original design is problematic from two perspectives - separation of external tools as presented in section II and needed time for execution for each generated program from generation. The main idea of the presented tool is to utilize genetic-based algorithms to generate every program contained in generation based on combinations of these recommendations from static and dynamic analysis. An increased number of generated recommendations would significantly increase the execution time of the evaluation step.

Newly placed dynamic analysis as shown in Figure 2 is executed after the generated of programs. This allows us to utilize produced recommendations for evaluation of programs and filter. One of such examples which demonstrates ability to utilize output from recommendations is memory allocation on GPU platform. Evaluation and filter can reduce number of further evaluated programs based on memory allocation, e.g. if multiple programs use local memory and some programs do not, we can filter out programs which utilize local memory.

Filter is responsible for final decision, if program can be optimized further or it is not possible to optimize it further because of time constraints etc.. Input for the filter are recommendations for every program from dynamic analysis. Based on recommendations, filter will decide, programs filtered out.

IV. Verification

Updated design of optimization flow in tool for automated optimization and parallelization in heliospheric field presented in section III should fix flaws that existed in original design. For verification, we decided to implement simple prototype of tool for automated optimization and parallelization in heliospheric field for GPU platform.

Prototype contained essential logic for parsing domain-specific language, logic for evaluating available platform, static analysis, generator of programs, dynamic analysis (such as calculation number of threads) and evaluation of programs for GPU platform.

For evaluation of functionality of tool, we decided to implement simple 1D backward-in-time model from Geliosphere in input domain-specific and compare execution times. Execution time of implementation of 1D backward-in-time model for 117 millions of testing particles in Geliosphere was 3.75 minutes, while execution time of generated code was 3.5 minutes. Proven difference between execution times is caused by a bit different approach to decomposition and logging, which is not present in generated code.

V. Further research directions

Verification of optimization flow proved that tool for automated optimization and parallelization in heliospheric field can useful and generate efficient as code as we did in Geliosphere.

In further research, we want to focus on various evaluation and dynamic analysis approaches. Original design suggested implementation of filter based on execution times. With increased number of types of static and dynamic analysis execution times for every program in generation can last hours. Usage of alternative approaches can significantly reduce number of mutations in generation, which should be executed and evaluated based on execution time.

VI. Acknowledgment

This publication was realized with support of the Operational Programme Integrated Infrastructure in frame of the project: Intelligent systems for UAV real-time operation and data processing, code ITMS2014+: 313011V422 and co-financed by the Europen Regional Development Found.

References

[1] R. Kappl, "Solarprop: Charge-sign dependent solar modulation for everyone," *Computer Physics Communications*, vol. 207, pp. 386–399, 2016.

[2] M. Solanik, P. Bobík, and J. Genči, "Geliosphere – GPU implementation of cosmic rays modulation models in heliosphere," in *25th Conference of Slovak Physicists*, 2021.

[3] V. Mykhailenko and P. Bobik, "Statistical error for cosmic rays modulation evaluated by sde backward in time method for 1d model," *Fluids*, vol. 7, no. 2, p. 46, 2022.

[4] M. Solanik, "Tool for automated optimization and parallelization in heliospheric field," in *22st Scientific Conference of Young Researchers*, 2022, pp. 150–151.

[5] J. R. Koza, "Genetic programming as a means for programming computers by natural selection," *Statistics and computing*, vol. 4, no. 2, pp. 87–112, 1994.

The way to the methodological choice of explainability and interpretability methods

[1]*Miroslava PAVLUSOVÁ (1st year)*
Supervisor: [2]Ján PARALIČ

[1,2]Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1]miroslava.pavlusova@tuke.sk, [2]jan.paralic@tuke.sk

***Abstract*—This article is devoted to the topic of explainability and interpretability. Basic definitions and division of methods according to various criteria are mentioned. However, in order to create a high-quality methodology for choosing a suitable explainability method, it is not enough to know the specification of the methods, but it is also necessary to take into account other aspects, such as user role, domain and task type.**

***Keywords*—Explainability, explainable artificial intelligence, interpretability, machine learning**

I. Introduction

Artificial intelligence (AI) has affected many areas of our lives. Systems using AI have reached a level where almost no human intervention is required in their design and deployment, and they achieve better results than a human. Decisions derived from such systems can directly affect people, e.g. in medicine, transport or defence. The danger lies in creating and using decisions that are not justifiable, legitimate, or that simply do not allow obtaining a detailed explanation of their behavior [1]. There is a need to better understand the functioning of such models and their results. Improving the understanding of the system can also lead to correcting its shortcomings. Explainable artificial intelligence (XAI) has seen significant growth over the past few years precisely because of the need for greater acceptance of these systems by humans.

Researchers often use the terms interpretability and explainability interchangeably. But there is also a group of researchers who identify their differences and distinguish between these two terms. There is no specific mathematical definition of interpretability or explainability [2]. Miller defines interpretability as the degree to which one can understand the reason for a decision [3]. Explainability is described as the generation of post hoc explanations for black box models [1].

In recent years, several methods have been proposed to explain machine learning models [1], [2]. The differences between these methods are in the input data, the output, and the way these explanations are generated, so they can be more or less suitable for a particular use case and user. The aim of this article is to outline a suitable methodology for selecting explainability methods.

The article is structured as follows. The next section describes the importance of explainability. Then section 3 describes the types of explainability methods from a technical point of view. The fourth part is devoted to different approaches to the methodology of method selection. The final fifth chapter describes our approach to research and design of a new methodology.

II. Needs of Explainability

There is a trade-off between the accuracy of a machine learning model and its ability to produce explainable and interpretable predictions. On the one hand, there are black box models that include deep learning [3] and ensembles [4]. Deep neural networks contain hundreds of layers and millions of parameters, making them considered complex models. On the other hand, there are white box or glass box models. These provide interpretable results - with common examples including linear models [5] and decision tree-based models [6]. Although these models are better explainable and interpretable, they are not as powerful and do not achieve comparable accuracy compared to the previous ones. Fig. 1 shows the trade-off between accuracy and interpretability of machine learning algorithms.

Users often doubt the decisions suggested by systems using machine learning. In some cases, they oppose the decision of the model because their inference mechanisms are opaque, non-intuitive and incomprehensible to humans.

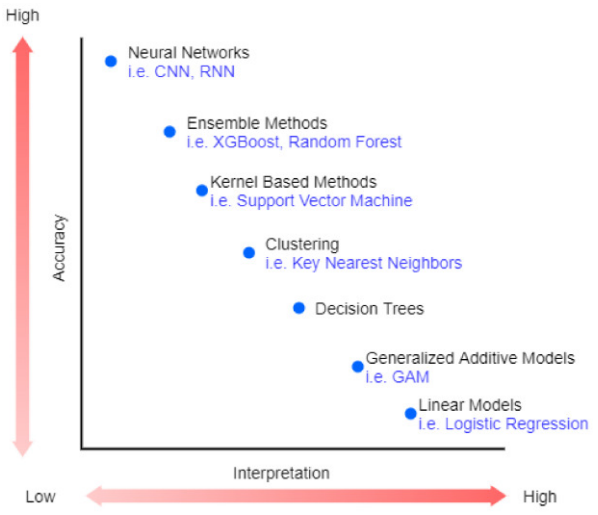


Fig. 1 The trade-off between accuracy and interpretability according to the characteristics of individual models [7].

The importance of explainability depends on the particular use case in the application domain. Adadi and Berrada [8] summarized the articles and formulated four motivations for explainability:

1. to justify decisions and respect the "right to explanation",
2. to allow the user control by identifying and correcting errors,
3. to help improve models by knowing why a certain output was produced, and
4. to gain new knowledge by examining learned prediction strategies.

Doshi-Velez and Kim [9] identified requirements in their research that can be optimized through interpretability and explainability:

- Fairness – it aims to make predictions unbiased and not discriminate against any group of people.
- Privacy – ensure that sensitive information is protected.
- Reliability/Robustness – guarantee that small changes in the input data do not cause large changes in the prediction.
- Causality - ensure that only causal relationships are picked up.
- Trust – it is easier for humans to trust a system that explains its decisions rather than a black box that just outputs the decision itself.

In the research [10], the authors expanded the list of user requirements that can be fulfilled using explainability methodss, e.g. acceptance, confidence, debugability, legal compliance, usability and others.

III. Technical Taxonomy of Interpretability Methods

There are several criteria according to which explanation methods and techniques can be classified from a technical point of view. A summary of this division is shown in Fig. 2.

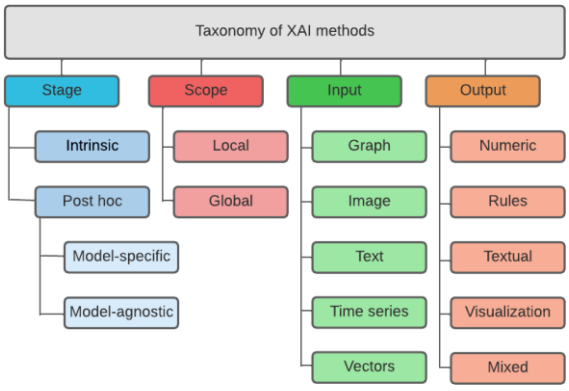


Fig. 2 Taxonomy of methods from technical view

A. Intrinsic vs. Post hoc

The first criterion is intrinsic vs. post hoc. This criterion distinguishes whether the interpretability is achieved through restrictions placed on the complexity of the model (intrinsic) or if it is a method that analyzes the model only after training (post hoc) [11].

Simple models that are interpretable due to their simple structure, such as decision trees or linear models belong to the intrinsic group. Internal interpretability is also defined as the transparency of the model and describes how the model works [12].

Post hoc interpretability refers to explanatory methods that are applied after the model has been trained. Post hoc methods include techniques used to convert an uninterpretable model into an explainable one.

B. Model-specific vs. model-agnostic

Model-specific interpretation methods are limited to specific models. Each method is based on the internal properties of the model.

Model-agnostic methods are applicable to any machine learning model after it has been trained. These methods analyze input and output pairs of models. These methods do not have access to the inner workings of the model.

C. Local vs. global

If a method explains an individual prediction, it is a local method. They usually approximate the behavior of the model around the instance the user wants to explain in order to get information about how the model works.

Global methods describe the average behavior of the entire model. These methods are often expressed as expected values based on the distribution of the data. They are especially useful when we want to debug the model.

D. Input and output data

Along with the main concepts, stages and extents of explainability, input and output formats have also been found to be important in the development of XAI methods [13] – [15]. Explainable methods differ when learning different types of input data such as images, numbers, texts, etc. At the same time, the different forms of explanation vary depending on the circumstances and expertise of the end users. The most common forms of explanations are numerical, rules, textual, visual and mixed.

IV. Different Methodologies of Explainable AI

With the increase in popularity of explainable methods, their number also increases and it is more difficult to navigate in them. How to find out which method to choose for a specific task? Currently, several solutions have been offered by experts. The basic pillar is the knowledge of technical taxonomy described in the previous section.

The taxonomy of existing machine learning interpretability methods proposed by Linardatos, Papastefanopoulos and Kotsiantis [16] for a multi-perspective comparison identifies four main categories for interpretability methods: methods to explain complex black-box models, methods to build white-box models, methods that promote fairness and limit the existence of discrimination, and finally methods for sensitivity analysis model predictions. Interpretation methods to explain deep learning models mostly apply to image classification and create saliency maps that highlight the influence of different image regions. The second group of methods that create white box models includes, in addition to linear models, decision trees and rule-based models, other more complex and sophisticated models that are equally transparent. Fairness is a sub-domain of machine learning interpretability that focuses solely on the social and ethical impact of machine learning algorithms by evaluating them in terms impartiality and discrimination. The study of fairness in machine learning is becoming broader and more diverse. The last category is

sensitivity analysis, which has seen an increase in the last few years. Numerous methods have been developed to create counterexamples, some of which focus on a more general environment, while others are tailored to specific types of data. The disadvantage of the proposed taxonomy is its focus only from the point of view of models.

Map of explainable approaches was was created as another approach [17]. Authors have undertaken a survey to help industry practitioners and data scientists understand the field of explainable machine learning better and apply the right tools. In addition to the division of models into interpretable and opaque, they also addressed explainability categories and explainability principles, which are listed in Table 1. Looking at the different kinds of explanations, it is clear that each of them concerns a different aspect of explainability. This means that there is no one-size-fits-all approach. The aspects of the category of explainability and the principles of explainability are interesting in this solution. However, this taxonomy does not take into account the role of the user at all.

TABLE 1

EXPLAINABLE CATEGORIES AND PRINCIPLES

Explainable category	Explainable principle
Explanation by simplification	Rule-based learner Decision tree
Feature relevance explainations	Influence functions Sensitivity Game theory inspired Interaction based
Local explanations	Rule-based learner Linear approximation Counterfactuals
Visual explanations	Sensitivity Dependency plots
Explanation by simplification	Rule-based learner Decision trees/prototypes Distillation
Feature relevance	Feature importance

The authors of the study [18] describe that it is necessary to create a methodology that would allow mapping the needs of users in explainability of machine learning models and the properties of individual methods. Based on previous work, they define a template that can be used to document both aspects. Different methods are introduced to elicit needs from stakeholders in a particular explainability use case. Most of them use an approach similar to the development of information systems and software development. Stakeholders can be categorized by their roles in the organization [19], their experience with machine learning [20], or a combination of the two [21]. We could also categorize the needs of interested parties. Some authors mention possible high-level explainability goals, such as model tuning, monitoring, etc. Other provide a list of more detailed needs such as privacy, fairness, legal compliance, etc. After documenting the needs of the stakeholders, it is necessary to confront them with the knowledge base of the detailed characteristics of the explanation method to select the most appropriate explanation methods. This approach represents an important principle of linking needs and concrete methods. It could also be enriched with other aspects such as type of tasks.

V. OUR ONGOING AND FUTURE RESEARCH

The authors of the research [22] focused on the analysis of various articles describing the use of XAI from a domain and task perspective. Fig. 3 shows the distribution of application domains for the 137 articles included in this research.

Obviously, most XAI methods are developed to be domain agnostic. However, the largest use of XAI is in healthcare, which may be due to the requirement for an explanation from the end user's perspective. In healthcare, medical practitioners embrace evidence-based practice as the guiding principle, which combines the most up-to-date research with clinical knowledge and patient conditions. Implementing a non-interpretable machine learning model in medicine raises legal and ethical issues [23].

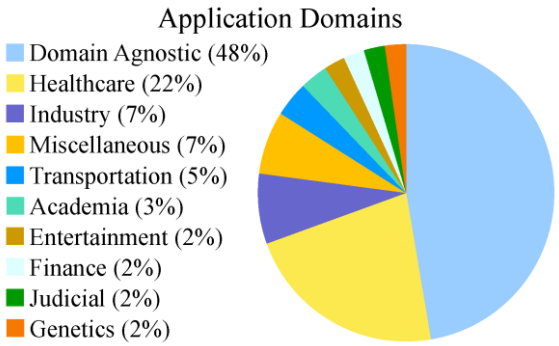


Fig. 4 Use of explainability methods in different domains [22]

It is clear that AI and ML methods are used for decision support systems in many application domains, and the need for XAI is high, as can be seen in Fig. 4, which describes the application tasks that occurred in the examined articles. application tasks

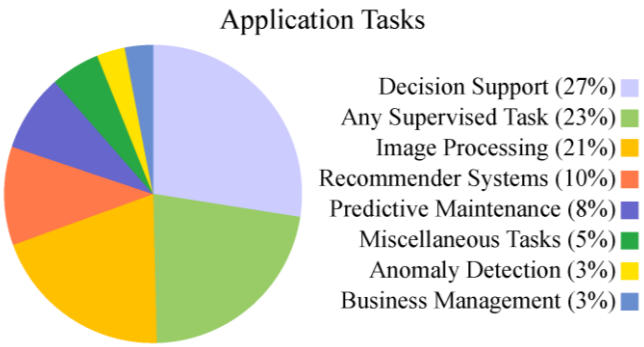


Fig. 3 Different application task for XAI methods [22]

The goal of our future research will be to propose a specific methodology that can guide in solving the explainability use case. The created methodology should include various aspects. In principle, each use case has three main parts: Who? What? How?

- Who - focus on the type of user, his knowledge, limitations, domain, form of explanation,

applicability of explanation, etc.
- What - user needs, task type.
- How - technical parameters of methods, data types, evaluation.

The resulting methodology must also be comprehensible and will be tested in several domains such as medicine, finance and social media.

ACKNOWLEDGMENT

This work was partially supported by the Slovak Grant Agency of the Ministry of Education and Academy of Science of the Slovak Republic under grant no. 1/0685/21.

REFERENCES

[1] A. Barredo Arrieta *et al.*, "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Information Fusion*, vol. 58, pp. 82–115, Jun. 2020, doi: 10.1016/J.INFFUS.2019.12.012.

[2] C. Molnar, *Interpretable Machine Learning*. 2022. Accessed: Feb. 20, 2023. [Online]. Available: https://christophm.github.io/interpretable-ml-book/cite.html

[3] T. Miller, "Explanation in artificial intelligence: Insights from the social sciences," *Artif Intell*, vol. 267, pp. 1–38, Feb. 2019, doi: 10.1016/J.ARTINT.2018.07.007.

[4] M. A. Ganaie, M. Hu, A. K. Malik, M. Tanveer, and P. N. Suganthan, "Ensemble deep learning: A review," *Eng Appl Artif Intell*, vol. 115, Apr. 2021, doi: 10.1016/j.engappai.2022.105151.W.-K. Chen, *Linear Networks and Systems* (Book style). Belmont, CA: Wadsworth, 1993, pp. 123–135.

[5] S. Weisberg, "Applied linear regression," 2005, Accessed: Feb. 20, 2023. [Online]. Available: https://books.google.com/books?hl=sk&lr=&id=xd0tNdFOOjcC&oi=fnd&pg=PR7&ots=dV7uypIBIQ&sig=d7JqmG3pfbVxdloPUokHmU_aLRo

[6] S. R. Safavian and D. Landgrebe, "A Survey of Decision Tree Classifier Methodology," *IEEE Trans Syst Man Cybern*, vol. 21, no. 3, pp. 660–674, 1991, doi: 10.1109/21.97458.

[7] T. A. A. Abdullah, M. S. M. Zahid, and W. Ali, "A Review of Interpretable ML in Healthcare: Taxonomy, Applications, Challenges, and Future Directions," *Symmetry 2021, Vol. 13, Page 2439*, vol. 13, no. 12, p. 2439, Dec. 2021, doi: 10.3390/SYM13122439.

[8] A. Adadi and M. Berrada, "Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)," *IEEE Access*, vol. 6, pp. 52138–52160, Sep. 2018, doi: 10.1109/ACCESS.2018.2870052.

[9] F. Doshi-Velez and B. Kim, "Towards A Rigorous Science of Interpretable Machine Learning," Feb. 2017, doi: 10.48550/arxiv.1702.08608.

[10] M. Langer *et al.*, "What do we want from Explainable Artificial Intelligence (XAI)? – A stakeholder perspective on XAI and a conceptual model guiding interdisciplinary XAI research," *Artif Intell*, vol. 296, p. 103473, Jul. 2021, doi: 10.1016/J.ARTINT.2021.103473.

[11] D. v. Carvalho, E. M. Pereira, and J. S. Cardoso, "Machine Learning Interpretability: A Survey on Methods and Metrics," *Electronics 2019, Vol. 8, Page 832*, vol. 8, no. 8, p. 832, Jul. 2019, doi: 10.3390/ELECTRONICS8080832.

[12] Z. C. Lipton, "The mythos of model interpretability," *Commun ACM*, vol. 61, no. 10, pp. 35–43, Oct. 2018, doi: 10.1145/3233231.

[13] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi, "A survey of methods for explaining black box models," *dl.acm.org*, vol. 51, no. 93, Aug. 2018, doi: 10.1145/3236009.

[14] S. Wachter, B. Mittelstadt, and C. Russell, "Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR," *SSRN Electronic Journal*, Nov. 2017, doi: 10.2139/SSRN.3063289.

[15] G. Vilone, L. L.-M. L. and K. Extraction, and undefined 2021, "Classification of explainable artificial intelligence methods through their output formats," *mdpi.com*, vol. 3, no. 3, pp. 615–661, Sep. 2021, doi: 10.3390/make3030032.

[16] P. Linardatos, V. Papastefanopoulos, and S. Kotsiantis, "Explainable AI: A Review of Machine Learning Interpretability Methods," *Entropy 2021, Vol. 23, Page 18*, vol. 23, no. 1, p. 18, Dec. 2020, doi: 10.3390/E23010018.

[17] V. Belle and I. Papantonis, "Principles and Practice of Explainable Machine Learning," *Front Big Data*, vol. 4, p. 39, Jul. 2021, doi: 10.3389/FDATA.2021.688969/BIBTEX.

[18] T. Vermeire, T. Laugel, X. Renard, D. Martens, and M. Detyniecki, "How to Choose an Explainability Method? Towards a Methodical Implementation of XAI in Practice," *Communications in Computer and Information Science*, vol. 1524 CCIS, pp. 521–533, 2021, doi: 10.1007/978-3-030-93736-2_39/FIGURES/1.

[19] U. Bhatt *et al.*, "Explainable Machine Learning in Deployment," *FAT* 2020 - Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pp. 648–657, Sep. 2019, doi: 10.48550/arxiv.1909.06342.

[20] R. Yu, L. S.-V. Informatics, and undefined 2018, "A user-based taxonomy for deep learning visualization," *Elsevier*, Accessed: Feb. 20, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2468502X1830038X

[21] H. Suresh, S. R. Gomez, K. K. Nam, and A. Satyanarayan, "Beyond expertise and roles: A framework to characterize the stakeholders of interpretable machine learning and their needs," *Conference on Human Factors in Computing Systems - Proceedings*, May 2021, doi: 10.1145/3411764.3445088.

[22] M. R. Islam, M. U. Ahmed, S. Barua, and S. Begum, "A Systematic Review of Explainable Artificial Intelligence in Terms of Different Application Domains and Tasks," *Applied Sciences 2022, Vol. 12, Page 1353*, vol. 12, no. 3, p. 1353, Jan. 2022, doi: 10.3390/APP12031353.

[23] J. Amann, A. Blasimme, E. Vayena, D. Frey, and V. I. Madai, "Explainability for artificial intelligence in healthcare: a multidisciplinary perspective," *BMC Med Inform Decis Mak*, vol. 20, no. 1, Dec. 2020, doi: 10.1186/S12911-020-01332-6.

Incorporating Visual Features of Documents for Improved Information Extraction in the Legal Domain

[1]*Tatiana KUCHČÁKOVÁ (2nd year),*
Supervisor: [2]*Jaroslav PORUBÄN*

[1,2]Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

[1]tatiana.kuchcakova@tuke.sk, [2]jaroslav.poruban@tuke.sk

***Abstract*—The extraction of relevant information from legal documents, such as contracts and court opinions, using natural language processing and machine learning techniques is an essential task in the legal domain. By incorporating visual features such as the structure of text and the type of font used, the system can gain valuable insights into the meaning and intent behind legal texts. The use of visual features can provide additional information that cannot be obtained from textual or semantic features alone, which can help the system to make better predictions or decisions. The paper describes initial experiments using machine learning techniques to identify entities in contracts and proposes a plan for incorporating visual features into Natural Language Processing (NLP) models.**

***Keywords*—visual features, document layout analysis, information extraction, named entity Recognition, legal processing**

I. Introduction

Legal domain information extraction refers to the process of automatically extracting relevant information from legal documents using natural language processing and machine learning techniques [1]. Information extraction in the legal domain typically involves identifying named entities like parties, dates, locations, and legal concepts, as well as extracting relationships between them, such as obligations, prohibitions, and rights [2]. This can be done by using named entity recognition or relationship extraction. Legal domain information extraction has a variety of practical applications, like contract analysis [3], risk assesment [4], and classification of contract-amendment relationships [5]. According to article [6], the inclusion of page features such as header and footer information, page numbers, and font styles during the preparation of legal documents can enhance the precision of NLP algorithms in categorizing and classifying text components. The reason behind this is that page features can provide additional context and information about the document's content, thus enabling NLP algorithms to more accurately comprehend the structure and significance of the text. Layout analysis in genral involves several tasks, including page segmentation, document structure analysis, and content extraction [7]. Document layout analysis can help improve information extraction by:

- Identifying regions of interest: By analyzing the layout of a document, we can identify regions that are likely to contain relevant information. This can help in directing the information extraction process to focus on the relevant parts of a document.
- Resolving ambiguities: Document layout analysis can help resolve ambiguities in legal documents by identifying the context in which certain terms or phrases appear.
- Handling tables and structured information: Tables and other structured information are common in legal documents, and document layout analysis can help in identifying the structure of such information.

The layout of a document can vary widely even within a single domain, such as legal contracts. This variability can be due to differences in the structure of the contract, the formatting of the text, or the use of different legal terminology or language.This variability can make it difficult to develop NLP models that can accurately extract information from contracts, since the models need to be able to recognize and understand the different layouts and formatting used in different contracts.

II. Initial experiments

Given that legal documents are often highly confidential, we have used a publicly available source of contracts from the Slovak Central Register of Contracts. Under the Slovak Law of Free Access to Information, obligated entities such as state authorities, municipalities, higher territorial units, as well as legal and natural persons with the power to decide on the rights and obligations in the field of public administration, are required to upload contracts to this central registry of contracts. Expanding on our earlier work [8], we have investigated several approaches for recognizing legal entities in contracts, such as employing rule-based entity labeling with regular expressions and machine learning methods. For our machine learning approach, we employed a Bi-Directional Long Short-Term Memory (Bi-LSTM) model in combination with a Conditional Random Field (CRF) layer. To prepare the dataset, we manually labeled 4316 sentences extracted from 112 lease contracts obtained from a registry. During the labeling process, we focused on two entities: lessor and lessee. We used the Inside-Outside-Beginning (IOB) format for tagging the entities in the text. With the labeled data, we trained our Bi-LSTM-CRF model. The trained model achieved an overall accuracy of 0.81, with varying results across different classes. For the 'B-P' class, which stands for the beginning of the lessor entity, the model achieved 0.60 precision, 0.55 recall, and 0.57 F1-score. For the 'B-N' class, which represents the beginning of

the lessee entity, the results were 0.75 precision, 0.68 recall, and 0.71 F1-score. The registry under examination comprises of two different types of PDF formats: textual and scanned. In cases where the scanned contracts are processed, Optical Character Recognition technology (OCR), such as Tesseract, is utilized to convert the visual text into machine-readable text. In these initial experiments, Tesseract 4.1.3 was used as the OCR engine, which uses LSTM-based technology. Prior to OCR, a preprocessing step of thresholding was performed on the input images using OpenCV's thresholding function. This was done to improve the OCR results by enhancing the contrast between the text and the background of the input images. However, this process is not without its flaws and resulted in the loss of crucial data due to the imperfect nature of OCR technology. Furthermore, standard NLP techniques that are applied to the plain text obtained from PDFs, do not account for visual features, such as text intention or font style, as well as the overall document layout and typography, that are vital for humans to comprehend the contract effectively. We postulate that incorporating these visual features into the entity recognition process could potentially improve the accuracy of the system.

After analyzing the existing research in the field, we have made the decision to enhance the identification of entities in contracts by incorporating visual characteristics of contract documents during the feature engineering stage of our NLP pipeline. However, this approach presents a number of challenges that must be addressed in our research plan.

III. Proposed plan

In order to put this plan into action, a potential strategy would involve utilizing computer vision methodologies to extract visual characteristics from contracts, alongside extracting machine-readable metadata from the contracts. The subsequent stage would be to represent these features in a format that can be leveraged by NLP algorithms, and then merge them with NLP models to facilitate information extraction. This can be accomplished through the inclusion of extracted visual features as supplementary input characteristics to NLP models, or by integrating computer vision and NLP models into a unified deep learning architecture.

Fig. 1 shows a possible pipeline for case when a contract is in textual PDF format. In this case PDF streams can be used to obtain visual features during the feature extraction stage. PDF stream is an object within the PDF file that contains binary data. These streams are compressed to reduce file size, and they need to be decoded and processed during PDF parsing. PDF parsing libraries handle PDF streams by decoding and extracting the relevant information [9]. This can include text, images, and metadata, as well as other elements such as fonts, colors, and document structure. Feature extraction is then followed by incorporating these visual features into the NLP model, to possibly improve the accuracy when identifying legal entities within contracts.

Fig. 2 illustrates a potential pipeline for processing contracts in image format. In this pipeline, we leverage all visual elements of the contract through image processing techniques. Computer vision can be used to segment regions of interest containing key entities based on labeled images. This approach is suitable for both scanned documents and text converted to images. It is important to note that, particularly for scanned

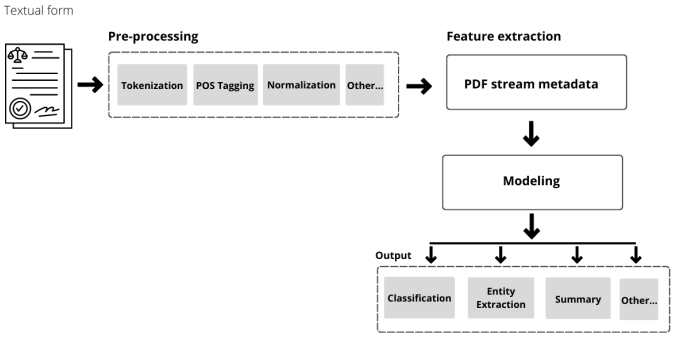


Fig. 1. NLP pipeline for incorporating visual features from PDFs in textual form

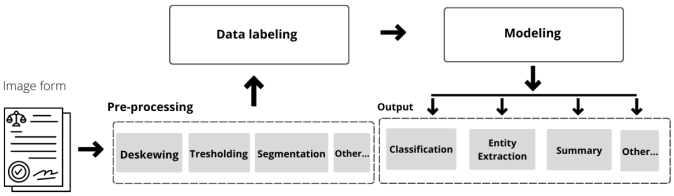


Fig. 2. Pipeline for extracting visual features in a way of image processing

documents, pre-processing is crucial to address issues such as color imbalance, document rotation, and other similar errors in the image processing stage.

It is possible to combine both approaches into a single robust multimodal system. This unified pipeline can handle a wider range of contract information and leverage the strengths of both PDF parsing and computer vision techniques. Incorporating the extracted features from both pipelines into a single NLP model which can be designed to handle the diverse set of features can improve the accuracy when identifying legal entities within contracts.

Acknowledgment

This work was supported by project VEGA No. 1/0630/22 "Lowering Programmers' Cognitive Load Using Context-Dependent Dialogs".

References

[1] I. Chalkidis, I. Androutsopoulos, and A. Michos, "Extracting contract elements," in *Proceedings of the 16th edition of the International Conference on Articial Intelligence and Law*, 2017, pp. 19–28.

[2] D. Hendrycks, C. Burns, A. Chen, and S. Ball, "Cuad: An expert-annotated nlp dataset for legal contract review," *arXiv preprint arXiv:2103.06268*, 2021.

[3] R. Dale, "Law and word order: Nlp in legal tech," *Natural Language Engineering*, vol. 25, no. 1, pp. 211–217, 2019.

[4] G. Baader and H. Krcmar, "Reducing false positives in fraud detection: Combining the red flag approach with process mining," *International Journal of Accounting Information Systems*, vol. 31, pp. 1–16, 2018.

[5] F. Song, "Classification of contract-amendment relationships," *arXiv preprint arXiv:2106.14619*, 2021.

[6] F. Josi, C. Wartena, and U. Heid, "Preparing legal documents for nlp analysis: Improving the classification of text elements by using page features," in *Computer Science & Information Technology (CS & IT)*. AIRCC Publishing Corporation, 2022, pp. 17–29.

[7] V. Arlazarov, E. Andreeva, K. Bulatov, D. Nikolaev, O. Petrova, B. Savelev, and O. Slavin, "Document image analysis and recognition: a survey," *Computer Optics*, vol. 46, no. 4, p. 567 – 589, 2022.

[8] T. Kuchčáková and J. Porubän, "Processing legal contracts using natural language processing techniques," in *Proceedings of the 22nd Scientific Conference of Young Researchers*, 2022, pp. 226–229.

[9] M. S. U. Miah, J. Sulaiman, T. B. Sarwar, A. Naseer, F. Ashraf, K. Z. Zamli, and R. Jose, "Sentence boundary extraction from scientific literature of electric double layer capacitor domain: Tools and techniques," *Applied Sciences*, vol. 12, no. 3, 2022.

Feasibility of adding renewable energy, storage system and Electric Vehicle Charging Infrastructure in the Industrial zone

[1]*Rikin TAILOR (3rd year)*
Supervisor: [2]Ľubomír BEŇA

[1,2]Dept. of Electrical Power Engineering, FEI TU of Košice, Slovak Republic

[1]rikin.jitendrakumar.tailor@tuke.sk, [2]lubomir.bena@tuke.sk

***Abstract*—The demand for goods and services are increased rapidly in the past decade. To cope up with this demand, the industry has increased its production line, and similarly, the service industry has done a sizeable increment. This has created an immense burden on electrical utilities. However, another challenge for the utility industry is the new 2 pillars: digitization and climate change. A new phase of vehicles is also obliged to follow both pillars. Because of that many governments have undertaken in increasing the usage of EVs and their infrastructure. To support the overall demand for electricity and EVs charging infrastructure we have examined a couple of research papers and did research to understand the feasibility of renewable energy, battery storage system, and EVs charging infrastructure in the present grid.**

***Keywords*—digitization, Industrial distribution storage, renewable energy generation, Electric Vehicle (EVs), EVs infrastructure.**

I. INTRODUCTION

Industries are in a phase of a new revolution, as per industry 4.0. Industry 4.0 has replaced many manuals works with automation and robotics driven. Telecommunication network 5G has enabled Future Cloud Infrastructure (FCI), digitization, and quantum computing. Not only this but also all manufacturing industry, service industry, and business profiles become swift and smooth. At the same time, the requirements for goods and services are also rising with the rise in population and their demand. New technology and demand simultaneously create huge challenges for the electrical industry to generate and supply electricity.

Adding to it, the electrical industry is bound to adopt renewable energy generation due to the climate change resolution. Similarly, EV infrastructure is also on the verge of completely reforming the transportation industry. This cause further increases in electricity need. The task is not easy to fulfill this demand. As a consequence of which, consumers will have to participate in the distributed generation and storage facilities. Many governments are giving subsidiaries for installing rooftop photovoltaic at residential location usage of electric vehicles. However, research claims that distribution is very easily get damaged due to the random charging cycles of EVs and intermittent generation of grid-connected solar panels [1]. Compared to the residential distribution grid, the industrial grid is much more stable in operation. In the case of an electric vehicle charging system, the USA's almost 73% of electric vehicle charging electricity can be fulfilled by the HV network [2]. According to Hardman et al, 15-20% of charging incidents are observed at the industrial network [3]. It can cover more with predictable expansion. This made me think to check the practicability of the industrial networks to include renewable generation, storage systems and EVs charging infrastructure.

II. INDUSTRIAL INFRASTRUCTURE AND RESEARCH

Industries work on very high-power ratings. Modern Industries are also designed in such a way that it is possible to have predictable expansion in their growth. Here the discussion topic is the readiness of the industrial grid for the upcoming big transformation towards renewable energy, zone-wise storage system, and electric vehicle charging infrastructure. In the technical term, we call it a smart grid.

The understanding of modeling, monitoring, and management of power dynamics in industrial grid systems is given by Nina F. in her research paper [4]. With an increase in industrial load, it is also a challenging task to do a loaded schedule. Research based on the generalized load scheduling by understanding the interdependencies of load sequence is presented in [5], and the well-structured framework is introduced for all kinds of industries.

Many industries are ready to accept new changes and aim to increase their dependencies on renewable generation at their own premises. A similar concept can be seen in one of Turkey's biggest refinery plants, refinery has 5 MW of storage plant and 10 MW of its own wind generation capacity [6]. One more big storage capacity system is available in Shenhua's Wuhai coal mine site, which has a capacity of 3MW/2MWH [7].

Mansour Alramlawi has demonstrated in his paper that industrial load under blackout conditions can also withstand with optimal support of PV, Diesel generators and battery storage systems [8]. Optimize methods are performed to minimize the power input from the diesel generator and maximize the utilization of PV.

As we are advancing towards the electric transportation system, more and more electric charging stations are also required in all stages (Residential charging, Commercial charging, public charging). The impact of this charging station in the present grid in the Czech Republic is given in the article [9].

Many smart strategies are also introduced for charging electric vehicles in existing infrastructure [10,11,12]. The technical readiness of the industrial grid is quite high. It is also needed to consider that random charging methods of vehicles can disturb the balance of the industrial grid. However, if it is planned in an optimal way then it can serve several benefits as well. A comparable concept is introduced by Sheeraz Iqbal, where he talked about the G2V as well as the V2G concept and gave scenarios that how this technique adds the feature to the industrial grid for primary frequency control and congestion management [13]. The concept of finding the best possible location for electric vehicle charging infrastructure in an Industrial zone is possible with the genetic algorithm approach [14].

To protect sensitive equipment in the industry, grid power quality must be very well practiced. Supportive information, where the stabilizing device is introduced to improvise the power quality and stability for the industrial as well as medium voltage grid [15].

III. Distribution Grid Understanding

We have tried to design a solution for the distribution (Low voltage) level before applying the concept to the medium and industrial level solution with renewables, grid storage and electric vehicle charging infrastructure.

A. Electrical Energy Flow Algorithm for Household, Street and 2 Battery charging in Smart Street Development [16]

In this journal, we have simulated the street consisting of 6 houses with different loads, different PV generation and different charging cycles for the electrical vehicle. The purpose of this journal was to understand the behavior of low-voltage grids while applying smart distribution strategies. The overall concept was presented with the Smart control panel for the house as well as the central distribution stage, 3 algorithms are introduced which help to take action for electricity as well as communication flow and maintain the stability in the grid.

We have verified the Smart distribution structure with the 2 different cases of the normal grid and Faulty grid. In both cases, the system has responded very well, and we manage to run the system independently from the utility grid even in peak hours.

B. Design of Management Systems for Smart Grid [17]

The article is designed to demonstrate 2-way energy and communication flow with the peer-to-peer exchange. There is a smart control panel between the consumer and utility connection which enables the consumer to participate actively in the energy market with renewable generation at the distribution side. Similar to a journal, in this management system also algorithm, is introduced and tested with two different case studies. The system has performed quite well with the designed structure and helps the user with their rights to use electricity on their own time. It has also benefited users to use electricity in peak time with a full load and cost saving.

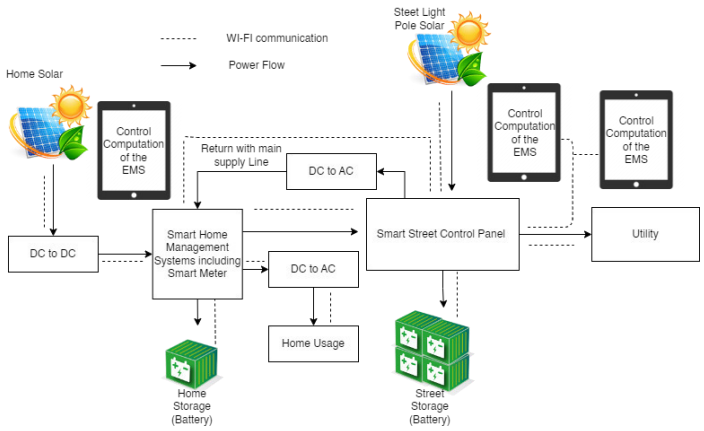


Fig. 1. Architecture Diagram of Power and Communication Flow [17]

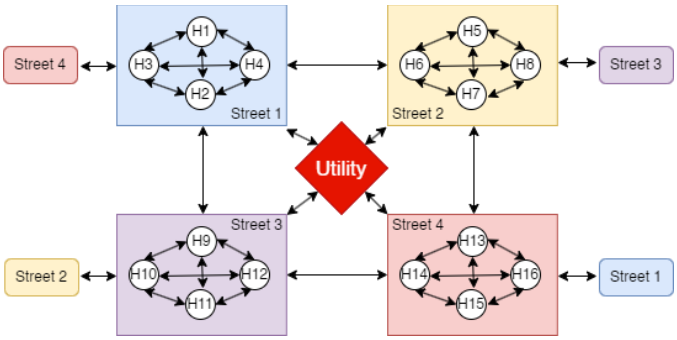


Fig. 2. Power exchange in wide network [17]

Figure 1 show the architecture diagram used for power and communication flow in smart distribution management system. Figure 2 indicates peer-to-peer energy transfer as well exchange of electricity between two zones.

IV. Future Scope

After understanding the overall industrial grid structure, and feasibility of expansion, the next steps are to create a complete simulation in MATLAB for the industrial zone by taking actual data of industries and adding a practical possible extension of PV generation as well storage units both zone and individual. This concept will be served with individual and zone-wise electrical charging infrastructure.

I will also try to demonstrate the possibility of EV charging infrastructure in a present industrial grid with real-time data.

Acknowledgment

This work was supported by the Slovak Research and Development Agency under contract No. APVV-19-0576, No. APVV-21-0312 and the Ministry of Education, Science, Research and Sport of the Slovak Republic and the Slovak Academy of Sciences under contract No. VEGA 1/0757/21.

References

[1] Gong X, Lin T, Su B H, Impact of Plug-in Hybrid Electric Vehicle Charging on Power Distribution Network [J]. Power System Technology, 2012, 36(11): 30-35

[2] Richardson P, Flynn D, Keane A., Impact assessment of varying penetrations of electric vehicles on low voltage distribution systems[C]//IEEE Power and Energy Society General Meeting. Minneapolis: IEEE 2010: 1-6

[3] Hardman, S., et al., 2018. A review of consumer preferences of and interactions with electric vehicle charging infrastructure. Transp. Res. Part D: Transp. Environ. https://doi.org/10.1016/j.trd.2018.04.002

[4] Nina F. Thornhill, Davide Fabozzi, Bikash C. Pal, Monitoring and Management of Power Transmission Dynamics in an Industrial Smart Grid, The financial support from the Marie Curie FP7-IAPP project "Using real-time measurements for monitoring and management of power transmission dynamics for the Smart Grid – REAL-SMART.

[5] Madhusudan Kumar, Mala De, Optimal Load Scheduling for Industrial Load – Analysis for a Generalized Industrial Load Model, 978-1-7281-4103-9/19/$31.00 ©2019 IEEE.

[6] Hasan Basri Cetinkaya, Bayram Kurucay, Panagiotis Stamoulis, Turan Turhan, Oktay Erisik, Case Study on the Use of Energy Storage in Industrial Plants with a Renewable Energy Plan, 2020 7th International Conference on Electrical and Electronics Engineering, 978-1-7281-6788-6/20/$31.00 ©2020 IEEE.

[7] George Zhou, Francis Wang, Tong Wu, Xiaodong Zhao, Shen Chen, Energy Storage Based Industrial Power

[8] Mansour Alramlawi, Aouss Gabash, Erfan Mohagheghi, Optimal Operation of PV-Battery-Diesel MicroGrid for Industrial Loads Under Grid Blackouts, 978-1-5386-5186-5/18/$31.00 c 2018 IEEE

[9] Oliver Marcincin, Zdenek Medvec, Petr Moldrik, The Impact of Electric Vehicles on Distribution Network, 978-1-5090-6406-9/17/$31.00 ©2017 IEEE

[10] Rikin Tailor, Ľubomír Beňa, Zsolt Čonka, Electric Vehicle Charging Management System in the residential streets for individual houses using a smart controller in a Present Grid, Electroenergetika 2022

[11] R. Mehta, D. Srinivasan, A. M. Khambadkone, Smart Charging Strategies for Optimal Integration of Plug-in Electric Vehicles within Existing Distribution System Infrastructure, DOI 10.1109/TSG.2016.2550559

[12] Lu Xia, Iven Mareels, Tansu Alpcan, Marcus Brazil, Julian de Hoog, Doreen A. Thomas, A Distributed Electric Vehicle Charging Management Algorithm Using Only Local Measurements, 978-1-4799-3653-3/14/$31.00 ©2014 IEEE

[13] Sheeraz Iqbal, Mishkat Ullah Jan, Mohamed Abdelkarim Abdelbaky, Haseeb Ur Rehman, Salman Salman, Syed Asad Abbas Rizvi, Muhammad Aurangzeb, Aggregation of EVs for Primary Frequency Control of an Industrial Microgrid by Implementing Grid Regulation & Charger Controller, DOI 10.1109/ACCESS.2020.3013762

[14] Dimitrios Efthymiou1 & Katerina Chrysostomou1 & Maria Morfoulaki1 & Georgia Aifantopoulou1, Electric vehicles charging infrastructure location: a genetic algorithm approach, Received: 9 June 2016 /Accepted: 11 April 2017 /Published online: 5 May 2017 # The Author(s) 2017. This article is an open access publication, DOI 10.1007/s12544-017-0239-7

[15] Arun Kannan, Thorsten Reimann, Diana Strauss-Mincu, Marvin Rolle, Christian Dresel, Ensuring Power Quality and Stability in Industrial and Medium Voltage Public Grids, 978-1-5386-4612-0/18/$31.00 ©2018 IEEE

[16] Rikin Tailor, Zsolt Čonka, Michal Kolcun, Ľubomír Beňa, Electrical Energy Flow Algorithm for Household, Street and Battery charging in Smart Street Development, https://doi.org/10.3390/en14133771

[17] Rikin Tailor, Ľubomír Beňa, Zsolt Čonka, Michal Kolcun, Design of Management Systems for Smart Grid, 978-1-6654-2164-5/21/$31.00 ©2021 IEEE

Author's index

SCYR 2023: 23rd Scientific Conference of Young Researchers

Proceedings from Conference

Published: Faculty of Electrical Engineering and Informatics
Technical University of Košice
Edition I, 196 pages
Number of CD Proceedings: 50 pieces

Editors: Prof. Ing. Alena Pietriková, CSc.
Assoc. Prof. Ing. Emília Pietriková, PhD.

ISBN 978-80-553-4377-8