ITIS 24
101101101100

Fakulteta za informacijske študije
Faculty of information studies

ITIS 2024 BOOK OF PROCEEDINGS

"Harnessing the power of big data for green and digital transition"

15TH INTERNATIONAL CONFERENCE ON INFORMATION TECHNOLOGIES AND INFORMATION SOCIETY

NOVEMBER 2024, LJUBLJANA, SLOVENIA

15th International Conference on Information Technologies and Information Society

ITIS 2024

Conference Proceedings

November 7-8, 2024

Ljubljana, Slovenia

Organized by

Fakulteta za informacijske študije
Faculty of information studies

Faculty of Information Studies in Novo mesto

Supported by

ARIS Slovenian Research and Innovation Agency

CONFERENCE COMMITTEES

Organizing Committee

Nuša Erman, *chair, Faculty of Information Studies in Novo mesto, Slovenia*
Biljana Mileva Boshkoska, *chair, Faculty of Information Studies in Novo mesto, Slovenia*
Blaž Rodič, *chair, Faculty of Information Studies in Novo mesto, Slovenia*
Aljaž Blatnik, *Faculty of Information Studies in Novo mesto, Slovenia*
Nika Robida, *Faculty of Information Studies in Novo mesto, Slovenia*
Kseniia Gromova, *Faculty of Information Studies in Novo mesto, Slovenia*
Mateja Lesar, *Faculty of Information Studies in Novo mesto, Slovenia*
Sanda Zakrajšek, *Faculty of Information Studies in Novo mesto, Slovenia*
Katja Peterlin, *Faculty of Information Studies in Novo mesto, Slovenia*
Alex Zorko, *Faculty of Information Studies in Novo mesto, Slovenia*

Program Committee

Zoran Levnajić, *chair, Faculty of Information Studies in Novo mesto, Slovenia*
Filipo Sharevski, *DePaul University, USA*
Sanda Martinčić – Ipšić, *University of Rijeka, Croatia*
Vesna Andova, *"SS. Cyril and Methodius" University in Skopje, North Macedonia*
Tomáš Hlavsa, *Czech University of Life Sciences Prague, Czech Republic*
Galia Marinova, *Technical University of Sofia, Bulgaria*
Jana Suklan, *Newcastle University, UK*
Małgorzata Pańkowska, *University of Economics in Katowice, Poland*
Marija Mitrović Dankulov, *Institute of Physics Belgrade, Serbia*
Dolores Modic, *Nord University, Norway*
Vida Vukašinović, *Jožef Stefan Institute, Slovenia*
Mateja Rek, *School of Advanced Social Studies, Slovenia*
Tea Golob, *Faculty of Information Studies in Novo mesto, Slovenia*
Srđan Škrbić, *Faculty of Information Studies in Novo mesto, Slovenia*
Mirjana Mikalački, *University of Novi Sad, Faculty of Science and Mathematics, Serbia*

Published by: Faculty of information studies in Novo mesto
Ljubljanska cesta 31a
8000 Novo mesto, Slovenia

Published in: 2024

Edited by: Nuša Erman

Kataložni zapis o publikaciji (CIP) pripravili v Narodni in univerzitetni knjižnici v Ljubljani / Cataloguing record of the publication (CIP) prepared by the National and University Library in Ljubljana

COBISS.SI-ID 215108355

ISBN 978-961-96549-1-0 (PDF)

EDITOR'S NOTE: Participants were asked to submit papers written in English. Keywords were chosen by participants and their number has not been reduced. Authors of submissions are responsible for the reliability of contents and other statements made in their work. Submissions are not proofread.

PROGRAM

	Thursday, 7 November	
8:30 - 9:00	*Registration, Meet & Greet*	
9:00 - 9:20	Welcome address by **Matej Makarovič, Dean of Faculty of Information Studies in Novo mesto**	
	Welcome address by **Mojca Štruc, Head of the Digital Inclusion Sector at the Ministry of Digital Transformation, Slovenia**	
9:20 - 10:00	**Keynote lecture by Dr. Francesco Tiezzi, University of Florence, Italy** **Unveiling Robotic Systems' Behaviour via Data Analysis**	**Session 1:** **Smart Cities and Industry** **Chair: Zoran Levnajić**
10:00 - 11:20	Pavle Boškoski and Vanja Subotić Statistical analysis of electrochemical impedance data	
	Jelena Joksimović, Jure Kos and Krištof Debeljak Comparative Analysis of Energy Consumption Prediction Models: LSTM, Interval-Based, and Hybrid Models	
	Andrej Trunkl and Pavle Boškoski Knowledge Transfer for Improved Detection of Electrical Insulators in Power Grid Photographs: Adapting Pre-trained Models to Slovenian Infrastructure	
	Valerij Grasic and Biljana Mileva Boshkoska Parameters devise for modelling real-time monitoring of magnetic field for the purpose of smart city using a mobile phone	
11:20 - 11:40	*Coffee break*	
11:40 - 12:20	**Keynote lecture by Aleš Simončič, Technical Sales Manager ISG, Lenovo SEE** **Sustainable HPC & AI Data Center – A Roadmap to CO2 Negative**	**Session 2:** **Blockchain and Business** **Chair: Mateja Lesar**
12:20 - 13:20	Dániel Kovács, Bálint Molnár and Viktor Weininger Enhancing Transparency and Efficiency in Small Size Loans Using Blockchain: A Shared Research Ledger Solution	
	Milica Stankovic and Gordana Mrdak Harnessing Big Data: Advantages and Challenges in Macroeconomic analysis and Economic Forecasting	
	Dániel Kovács, Bálint Molnár and Viktor Weininger Revolutionizing Pension Systems: The Integration of Blockchain with Business Process Management	
13:20 - 14:20	*Lunch break*	
14:20 - 15:00	**Keynote lecture by Dr. Ciara Heavin, University College Cork, Ireland** **Unlocking Big Data's Potential: Paving the Way for a sustainable Digital Future**	**Session 3:** **Healthcare and Biomedicine** **(Under the project AI2MED)** **Chair: Blaž Rodič / Biljana Mileva Boshkoska**
15:00 - 16:00	Mateja Lesar, Teja Štrempfel, Larisa Moderc, Lynn Süthoff, Jakob Sajovic, Peter Rogelj, Gorazd Drevenšek and Zoran Levnajić Can social exclusion be revealed from studying brain activity via EEG?	
	Mia Darkovska, Ilija Mizhimakoski, Vesna Andova, Jasmina Angelevska Kostadinoska and Ana Ristovska Dimitrovska Assessing Pigmented Skin Moles Using Minkowski Fractal Dimensions and Comprehensive Color Channel Analysis	
	Lana Dominkovic, Biljana Mileva Boshkoska and Aleksandra Rashkovska ECG-Derived Respiration on the Fantasia Dataset using the Signal Processing RRest Toolbox (zoom)	
16:00 - 16:20	*Coffee break*	
16:20 - 17:20	Vesna Stepišnik and Janez Kolar Factors of the development and manifestation of the talent of managers	**Session 4:** **Education and Talents** **(Under the project AI4VET4AI)** **Chair: Alenka Pandiloska Jurak**
	Laura Fink Application of prediction models in educational context	
	Alenka Pandiloska Jurak, Janez Kolar, Alan Berg, Stefan Mol, Leo Mršić, Scott Harrison and Gábor Kismihók Acquiring competencies for mastering tools that process big data - the case of AI usage skill gaps	
17:20 - 18:00	Ana Hafner and Dolores Modic Intellectual Property Rights Data for Sustainable Digitalisation and Digitalisation for Sustainability	**Session 5:** **Digital transfomation in society** **Chair: Alenka Pandiloska Jurak**
	Dorian Fildor Cluster Analysis of Key Links Between Digital Transformation and Sustainable Macroeconomic Growth in EU Countries	
18:20 - 19:00	*Social program*	
19:00 - 20:00	*Dinner*	
20:00 - 21:00	*Social program*	

	Friday, 8 November	
8:30 - 9:00	*Registration, Meet & Greet*	
9:20 - 10:00	**Keynote lecture by Dr. Franjo Cecelja, University of Surrey, United Kingdom** **Artificial Intelligence: A Circular Economy Perspective**	**Chair: Kseniia Gromova**
10:00 - 10:40	**Keynote lecture by Dr. Marko Debeljak, Institut Jožef Stefan, Slovenia** **AI-supported transition to sustainable agriculture**	
10:40 - 12:00	Jelena Topić Božič and Simon Muhič Data centres and sustainability: the effect of geographical location and electricity mix through life cycle approach	**Session 6:** **Trends in AI** **Chair: Kseniia Gromova**
	Mirjana Pejić Bach, Jasmina Pivar and Robert Fabac Emerging trends in generative artificial intelligence: Insights from patent analysis using Lens.org toolk	
	Bojan Pažek and Slavko Arh Topology-Driven Feature Extraction Based on PCA Pre-Alignment: Addressing Anisotropy in High-Dimensional Data (A Case Study)	
	Aljaž Blatnik and Blaž Rodič Modelling border crossing in irregular migrations on the western balkans route	
12:00- 12:20	*Coffee break*	
12:20 - 14:00	Jelena Joksimović Computer Vision for Time Series Classification	**Session 7:** **AI and NLP** **Chair: Aljaž Blatnik**
	Omkar Bihani, Janez Povh and Janez Zerovnik Towards a randomized algorithm for the maximum stable set problem	
	Fatima Aziz and Martin Žnidaršič Classification of bug severity with lexicon approaches	
	Ana Vladić and Juraj Petrović Attributing authorship to short texts on social networks	
	Ana Bezić Anthropology and Algorithms: The Good, the Bad, and Entangled	
14:00 - 15:00	*Break*	
15:00 - 15:45	Kseniia Gromova (JM Centre of Excellence TIA2030) and Cristian Gangaliuc Jean Monnet Centre of Excellence "Technology and Innovation for Agenda 2030 - EU Global Leadership - TIA2030" Round table discussion: Institutions, social networks and cognitive frames in implementation of Agenda 2030.	**Rountables**
15:45 - 16:30	Victor Cepoi (JM Module TIC203) and Anatolie Cosciug "A Market Constantly Changing: On the Return Migrant Entrepreneurs' Use of Human and Social Capital to Navigate Changes of the Operating Environment"	
16:30 - 17:00	Victor Cepoi Jean Monnet Module "Technology Innovation Communities 2030" Final considerations of the project.	

KEYNOTE SPEAKERS

Lectures by keynote speakers

Unveiling Robotic Systems' Behaviour via Data Analysis

by Francesco Tiezzi, *University of Florence, Italy*

Sustainable HPC & AI Data Center – A Roadmap to CO2 Negative

by Aleš Simončič, *Lenovo Data Center Group*

Unlocking Big Data's Potential: Paving the Way for a sustainable Digital Future

by Ciara Heavin, *University College Cork, Ireland*

Artificial Intelligence: A Circular Economy Perspective

by Franjo Cecelja, *University of Surrey, United Kingdom*

AI-supported transition to sustainable agriculture

by Marko Debeljak, *Institut Jožef Stefan, Slovenia*

CONFERENCE PAPERS

Statistical analysis of electrochemical impedance data

Pavle Boškoski[1,2], **Vanja Subotić**[3]
[1]Faculty of Information Studies, Ljubljanska cesta 31a, 8000 Novo mesto, Slovenia
[2]Jožef Stefan Institute, Jamova cesta 39, 1000 Ljubljana, Slovenia
[3]Institute of Thermal Engineering, Graz University of Technology
Inffeldgasse 25/B, 8010 Graz, Austria
pavle.boskoski@fis.unm.si, vanja.subotic@tugraz.at

Abstract. *This paper presents a comprehensive framework for the statistical analysis of electrochemical impedance spectroscopy data using Kullback- Leibler divergence, wavelet transforms, and variational Bayes inference. The proposed approach integrates discrete random binary signal excitation to enable faster and more efficient impedance measurements. It provides a comprehensive understanding of electrochemical systems such as lithium-ion batteries and solid-oxide fuel cell. The framework leverages statistical methods to allow for precise detection of changes in impedance characteristics and robust model selection. Experimental validation demonstrates the accuracy and reliability of the approach across a wide range of operating conditions. This makes it a valuable real-time diagnostics and monitoring tool in dynamic electrochemical environments.*

Keywords. solid-oxide fuel cell, Kullback- Leibler, variational Bayes, electrochemical impedance spectroscopy, discrete random binary signal

1 Introduction

Electrochemical impedance spectroscopy (EIS) is widely used for characterizing electrochemical energy systems, such as lithium-ion batteries and solid-oxide fuel cells (SOFCs). The critical advantage of EIS lies in its ability to provide valuable information about the internal state of these systems under various operating conditions. Traditionally, EIS measurements are performed by applying sinusoidal excitations at discrete frequencies. Still, this approach can be time-consuming, particularly in the low-frequency range, and is sensitive to environmental disturbances during prolonged measurement times [1, 2].

Recent advancements in impedance measurement techniques and model selection methods have significantly improved the accuracy and efficiency of EIS data analysis. This paper proposes a comprehensive statistical framework for analyzing impedance data using Bayesian inference, statistical divergence measures, and wavelet-based signal processing techniques. Our method aims to expedite impedance measurement and enhance the reliability of model selection through probabilistic approaches.

The conventional method for EIS analysis relies on mono- or multi-component sinusoidal signals, which limits the frequency resolution to only a few points per decade.

This leads to extended measurement times, especially in the low-frequency range, which is critical for assessing batteries' state of health (SOH) and state of charge (SOC). One promising approach to mitigate this issue is using broadband excitation signals, such as discrete random binary sequences (DRBS), which excite the system across a continuous frequency range. This technique enables faster impedance measurements by reducing the total experimental time while maintaining high-frequency resolution [1].

In parallel, methods for analyzing EIS data have evolved beyond traditional equivalent circuit models (ECMs). ECMs are helpful but often limited due to their reliance on a pre-determined model structure, which may not always capture the actual dynamics of fractional-order systems like SOFCs and lithium-ion batteries. The distribution of relaxation times has emerged as an alternative, offering a non-parametric method for analyzing impedance data [3]. However, both distribution of relaxation times and ECM suffer from model identifiability issues, where multiple models may provide similar fits to the data [4].

To address these challenges, Bayesian inference has been proposed for model selection in EIS analysis. By employing variational Bayesian (VB) inference, selecting the most probable model structure based on the observed data is possible while simultaneously estimating the model parameters [5]. This approach incorporates a probabilistic framework that selects the best-fitting model and quantifies uncertainty in both the model and the parameter estimates.

The combination of fast impedance measurement using discrete random binary signal (DRBS) and probabilistic model selection through Bayesian inference represents a significant improvement over existing techniques. This paper builds on these advances, proposing an integrated framework that applies statistical divergence measures for impedance data analysis and wavelet transforms for efficient frequency-domain conversion. Experimental validation of the proposed method on SOFCs and lithium-ion batteries demonstrates its efficacy in providing accurate and robust impedance analysis.

2 Impedance Measurement using DRBS and Wavelet Transform

Fast impedance measurement techniques have become a critical component in improving the efficiency of EIS data acquisition, especially when dealing with dynamic systems like lithium-ion batteries and SOFCs. This section explores how DRBS can be employed as an excitation signal for EIS measurements. We also introduce the statistical properties of DRBS and demonstrate how they can be exploited to assess the statistical properties of impedance data across a broad range of frequencies.

2.1 Statistical Properties of DRBS

A DRBS is a stochastic signal that switches between two values, typically $-a$ and $+a$, at random time intervals. The time between consecutive switches follows a Poisson distribution, which governs the signal's randomness and frequency characteristics. This random switching results in a broad frequency spectrum, making DRBS a highly effective excitation signal for systems that require a wide range of frequency inputs simultaneously [1].

The power spectral density of a DRBS, denoted as $S_{\text{DRBS}}(f)$, can be approximated as

a sinc function:

$$S_{\text{DRBS}}(f) = a^2\lambda\left(\frac{\sin(\pi f\lambda)}{\pi f\lambda}\right)^2$$

where λ is the minimum time interval between switches, and f represents the frequency. The spectrum of the DRBS is essentially flat over the range $[0, f_B]$, where $f_B = \frac{1}{2\lambda}$ is the bandwidth of the signal. This property makes the DRBS an ideal candidate for broadband excitation in EIS measurements, allowing simultaneous excitation of multiple frequency components [1].

The use of DRBS has several key advantages. It provides near-instantaneous excitation over a broad frequency range, enabling fast data acquisition. The signal's randomness ensures that the frequency components of the system response are uncorrelated, simplifying the statistical analysis of the impedance data. The stochastic nature of DRBS allows statistical techniques to assess the uncertainty and variability in the measured impedance.

2.2 Exploiting DRBS for Statistical Analysis of EIS

The statistical properties of DRBS make it particularly well-suited for assessing the statistical behavior of impedance in electrochemical systems. The impedance at a given frequency f can be calculated using the continuous wavelet transform (CWT) on the voltage $V(t)$ and current $I(t)$ signals:

$$Z(f) = \frac{W_V(t,f)}{W_I(t,f)} \tag{1}$$

where $W_V(t,f)$ and $W_I(t,f)$ are the wavelet coefficients of the voltage and current signals, respectively. By applying DRBS excitation, we obtain the impedance values at different frequencies and their statistical distributions over time. This enables us to calculate important statistical properties such as the mean, variance, and confidence intervals for the impedance at each frequency.

One of the most significant benefits of using DRBS excitation is the ability to estimate the probability distribution function of the impedance measurements. Given the random nature of the DRBS signal,the distribution of the impedance at each frequency f is [2]:

$$f_Z(z) = \frac{1-|\rho|^2}{\pi\sigma_u^2\sigma_i^2}\left(\frac{|z|^2}{\sigma_u^2} + \frac{1}{\sigma_i^2} - 2\frac{\rho_r z_r - \rho_i z_i}{\sigma_u\sigma_i}\right)^{-2}, \tag{2}$$

where z_r and z_i are real and imaginary components of the random variable Z. The location of the $f_Z(z)$ mode depends on the correlation coefficient ρ.

The distribution (2) has three parameters σ_u, σ_i, and the correlation coefficient ρ. These parameters can be easily estimated through the covariance matrix by using the calculated wavelet coefficients as

$$\begin{aligned} E\{Wu(t,f_0)Wu(t,f_0)^*\} &= \frac{\sigma_u^2}{2} \\ E\{Wi(t,f_0)Wi(t,f_0)^*\} &= \frac{\sigma_i^2}{2} \\ E\{Wi(t,f_0)Wu(t,f_0)^*\} &= \frac{\sigma_u\sigma_i}{2}\rho. \end{aligned} \tag{3}$$

2.3 Application to EIS

By exploiting the statistical properties of DRBS, we can obtain a more comprehensive understanding of the impedance characteristics of electrochemical systems. For instance, in lithium-ion batteries, variations in impedance across different frequencies can provide insight into the health of the battery's internal components, such as the anode, cathode, and electrolyte. Similarly, in SOFCs, changes in the impedance spectrum can be used to detect degradation processes and other performance-related issues.

The combination of DRBS excitation and wavelet-based impedance estimation offers a powerful tool for real-time monitoring and analysis of electrochemical systems. In addition to providing accurate impedance measurements, this approach enables a more robust statistical analysis of the data, allowing for better decision-making in applications such as battery management systems and fuel cell diagnostics.

3 Statistical Analysis of Impedance Data using Kullback-Leibler Divergence

The statistical properties of the DRBS introduced in the previous section provide a robust foundation for advanced statistical methods to analyze EIS data. In this section, we exploit these properties to calculate the Kullback- Leibler (KL) divergence, a measure that quantifies the difference in the impedance spectra between different operating conditions.

3.1 Kullback-Leibler Divergence for Impedance Analysis

The KL divergence is a statistical measure that quantifies the difference between two probability distributions. In the context of EIS, KL divergence allows us to measure how much the impedance spectrum under one set of operating conditions deviates from a reference condition. Given two probability distributions $p(Z(f))$ and $q(Z(f))$ representing impedance under two different conditions at frequency f, the KL divergence is defined as:

$$D_{\mathrm{KL}}(p \parallel q) = \int p(Z(f)) \log \left(\frac{p(Z(f))}{q(Z(f))} \right) dZ(f). \tag{4}$$

The KL divergence is non-negative and asymmetric, meaning $D_{\mathrm{KL}}(p \parallel q) \neq D_{\mathrm{KL}}(q \parallel p)$, and $D_{\mathrm{KL}}(p \parallel q) = 0$ only if $p = q$. This makes the KL divergence an effective tool for detecting changes in the impedance characteristics of electrochemical systems.

3.2 Application of KL Divergence to EIS Data

To apply KL divergence to EIS data, we first calculate the probability density functions $p(Z(f))$ and $q(Z(f))$ for the impedance at each frequency under two different operating conditions. These conditions may correspond to different states of charge (SOC) in a lithium-ion battery, varying operating temperatures in a fuel cell, or changes in electrolyte composition. For each frequency f, the KL divergence can be computed as:

$$D_{\mathrm{KL}}(p \parallel q) = \sum_{i=1}^{N} p(Z_i(f)) \log \left(\frac{p(Z_i(f))}{q(Z_i(f))} \right) \tag{5}$$

where $Z_i(f)$ represents the impedance values at frequency f, and N is the number of samples. This analysis allows us to identify the specific frequencies where the impedance

characteristics have changed most significantly, aiding in identifying degradation mechanisms or early signs of failure in electrochemical systems.

High values of KL divergence at specific frequencies indicate that the impedance distribution under the new operating condition significantly differs from the reference. For example, in a lithium-ion battery, an increase in KL divergence at low frequencies might indicate changes in charge transfer resistance or electrolyte degradation. In SOFCs, elevated KL divergence values could suggest performance degradation due to fuel composition changes or mechanical stresses on the cell. Thus, interpreting the KL divergence in EIS data provides valuable insights into the system's underlying physical processes and potential faults.

Using KL, divergence for EIS data analysis offers several advantages. KL divergence relies on probability distributions rather than model fitting, reducing the bias introduced by model-based analysis. Even small, statistically significant changes in impedance can be detected using KL divergence, making it useful for early detection of system degradation or failure. By exploiting the statistical properties of DRBS and the probability distributions of impedance data, KL divergence offers a sensitive and detailed method for analyzing changes in electrochemical systems, enhancing the diagnostic capabilities of EIS

3.3 Model Selection Using Statistical Properties of EIS with variational Bayes

The statistical properties of EIS data derived from DRBS excitation provide a natural framework for selecting the most appropriate model to describe the system's underlying dynamics. In complex electrochemical systems, such as lithium-ion batteries and SOFCs, the impedance behavior is often modeled using various ECMs. The challenge lies in choosing the best model while accounting for uncertainty and variability in the impedance data, which can be addressed effectively using variational Bayes (VB) inference.

By leveraging the probability distribution function (PDF) (2), the model selection problem becomes a probabilistic inference task. The goal is to identify the model M that best explains the observed statistical properties of the impedance data derived from EIS. This is where VB inference is compelling, as it balances model complexity with the ability to explain the data.

3.3.1 Variational Bayesian Inference for Model Selection

VB inference is an efficient method to approximate the posterior distribution of model parameters. It does this by maximizing the evidence lower bound (ELBO), which approximates the marginal likelihood of the data given a particular model. In the context of EIS, VB allows us to determine which model best fits the observed impedance distributions $p(Z(f))$ across a range of frequencies while also considering the uncertainty in the model parameters θ.

For a given set of candidate models $M_1, M_2, \ldots, M_k$, the posterior distribution of the model parameters θ, given the impedance data $Z(f)$, is approximated by a simpler, parameterized distribution $q(\theta)$. The ELBO for each model M is defined as:

$$\text{ELBO}(M) = \mathbb{E}_{q(\theta)} \left[\log p(Z(f)|\theta, M)\right] - \text{KL}(q(\theta) \parallel p(\theta|M)), \tag{6}$$

where $p(Z(f)|\theta, M)$ is the likelihood of the impedance data given the model parameters, and the second term is the KL divergence between the approximate posterior $q(\theta)$ and

the prior distribution $p(\theta|M)$. The VB inference process iteratively updates the approximate posterior distribution $q(\theta)$ to maximize the ELBO, thus selecting the model that provides the best explanation for the data while also avoiding overfitting. The stochastic nature of the impedance data, captured by the PDFs derived from DRBS, ensures that this model selection process accounts for both the mean behavior and the variability within the system.

3.3.2 Incorporating Statistical Properties for Model Selection

The statistical properties of the impedance data at each frequency play a crucial role in the VB model selection process. The likelihood term $p(Z(f)|\theta, M)$ depends on the impedance PDFs $f_Z(z)$ obtained from DRBS excitation. A model that accurately captures the shape and spread of these PDFs will have a higher likelihood, thus maximizing the ELBO. VB naturally incorporates uncertainty in the model parameters through the approximate posterior distribution $q(\theta)$. The stochastic impedance data derived from the DRBS provides a richer dataset for estimating this uncertainty, leading to a more robust and reliable model selection.

For instance, different ECMs may capture different physical phenomena, such as diffusion processes or charge transfer resistance. VB inference evaluates the ability of each model to explain the statistical properties of the impedance data across all frequencies. The model that most accurately captures the impedance distributions is selected by evaluating the ELBO for each candidate model.

Using VB, inference for model selection in EIS analysis offers several advantages. VB provides a systematic approach to incorporating parameter and model uncertainty, allowing for more confident model selection. By estimating the full posterior distribution of the model parameters, VB gives insight into the variability and uncertainty associated with each parameter, unlike traditional methods that provide point estimates. VB is computationally more efficient than other Bayesian methods, such as Markov Chain Monte Carlo, making it suitable for real-time or large-scale EIS data analysis.

By leveraging the statistical properties of impedance data derived from DRBS excitation, VB offers an efficient and robust method for model selection in EIS analysis. This approach ensures that the selected model fits the data accurately and reflects the physical processes governing the electrochemical system.

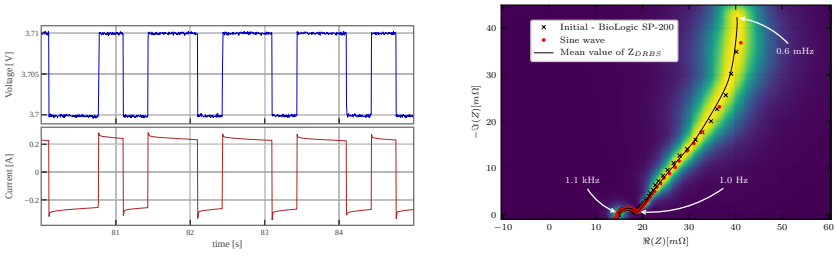
4 Experimental Validation

The proposed approach is validated on a solid-oxide fuel cell (SOFC) and a commercial lithium-ion battery. DRBS excitation and CWT are used to measure the impedance, and Bayesian inference is applied to select the optimal ECM for each system.

4.1 Battery impedance using DRBS excitation

Using DRBS, EIS was measured in the frequency region between 600 μHz and 1.1 kHz. The time domain signals of the excitation and response are shown in Figure 1a. EIS was measured at $\mathrm{V}_{DC} = 3.705$ V with peak-to-peak amplitude $\mathrm{V}_{AC} = 10$ mV.

The obtained EIS curve is shown in Figure 1b. The plot compares the initial impedance measurement performed with BioLogic SP-200, the impedance measured using DRBS signals, and the impedance measured using classical single-sine excitation signals. It can



(a) (b)

Figure 1: EIS measurement with DRBS on lithium-ion battery [1]. a) Time domain of the voltage excitation and the current response of the battery. b) Comparison of impedance calculated using DRBS and classical single-sine excitation signals..

be seen that the impedance between the two measurements is almost identical. Also, the value of the high-frequency resistance (R_S), where the imaginary part of the spectrum is zero, and the value of the impedance at low frequencies remains the same.

4.2 Application of Kullback- Leibler divergence on EIS data

For the case with low hydrogen content, the SOFC operated with 10% hydrogen in the inlet gas mixture. This phase lasted for 85 hours. As shown in Figure 2, there are no significant changes in the temperatures at both measurement points. An apparent initial decrease in the cell's voltage, particularly during the first 40 hours.

The Nyquist plots, shown in the Figure 2b, exhibit a clear trend from the initial measurement (red color) towards the end measurement (blue color). At first glance, the EIS curves exhibit changes throughout the observed frequency band. However, the most apparent change in the serial resistance is visible as a slight decay of the cell's voltage.

Figure 2b shows an apparent impedance change in the frequency range between 10 Hz - 4 kHz. Initially, after 40 hours of operation, the most intensive changes are in the frequency band between 200 – 800 Hz. After 70 hours of operation, a similar divergence change occurs at frequencies between 800 Hz – 4 kHz. Only slightly higher divergence values are observable in the 10 – 200 Hz frequency range. These changes can be linked to the shift of the impedance curve towards lower ohmic resistances from 49.5 to 48.5 mΩ and the behavior of the cell voltage, which mainly remained constant during the initial 20 hours but started to decrease from 1.32 to 1.31 V during ongoing operation. Generally, this can be interpreted as a new equilibrium slowly being achieved after the change in the inlet gas composition.

Additionally, Figure 2c shows increased impedance values scattered at frequencies < 1 Hz. These changes are also visible in the Nyquist diagram, whereby the low-frequency arc presents an unstable behavior. According to [6, 7, 8], this could be linked to changes in the diffusion losses, resulting from surplus H_2O which could still be present within the system from the previous experiment.

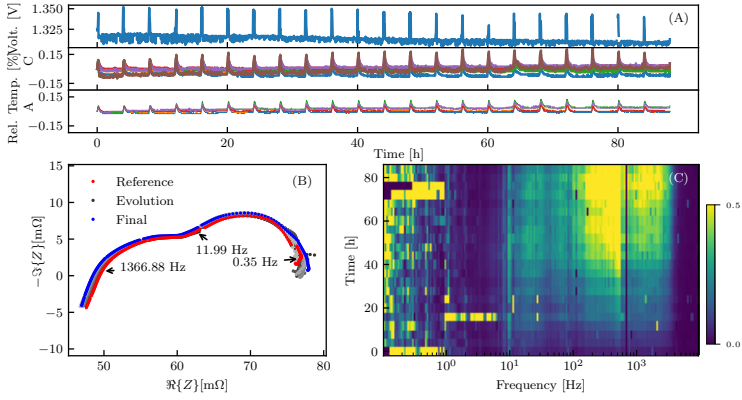


Figure 2: Results of operation under condition 2. Group (A) shows the time evolutions of the voltage and relative change of the temperatures at the anode-A and cathode-C sides. The relative change of the temperatures is calculated based on the initial measurement at $t = 0$. The local abrupt changes in the voltage and temperature values occur during the periodic EIS measurements. Plot (B) shows the evolution of the EIS curves from red to blue. Plot (C) shows the results of KL divergence evolution over the 85 hours during condition 2. The divergences are calculated to the reference (red) EIS curve. The most pronounced EIS changes are in the frequency band between 100 and 2000Hz.

4.3 SOFC experiment

The model selection approach was applied on a six-cell SOFC fuel cell stack running for 3600 hours [9]. The stack operated in an electric furnace at 750°C. Each cell had an active area of 80 cm^2. The fuel was a mixture of hydrogen and nitrogen with a ratio of 3/2 and flow of 0.36 Nl $\text{h}^{-1}\text{cm}^{-2}$. The air flow rate was 4 Nl $\text{h}^{-1}\text{cm}^{-2}$. At a nominal current of 32 A (0.4 A cm^{-2}) the stack operated with fuel utilisation of 77.5 %. During this period, the fuel cell went through three fuel utilisation cycles. The experiment ended when the ohmic loss in one of the cells reduced the stack's efficiency by 20%. The goal of the analysis here is to show the stability of the pole estimation approach and its sensitivity in cases where non-linearities are no longer negligible.

The top plot in Figure 3 illustrates the time evolution of the estimated number of poles M. The lower plot presents the root-mean-square error of the EIS model relative to the measured data. The root-mean-square error is calculated using the following equation

$$\begin{aligned}\text{RMSE} = &\sqrt{\frac{1}{S}\sum_{i}^{S}\left(\Re\{\hat{Z}(\omega_i)\} - \Re\left\{\text{E}[Z(\omega_i|\mathbf{z})]\right\}\right)^2} \\ &+ \sqrt{\frac{1}{S}\sum_{i}^{S}\left(\Im\{\hat{Z}(\omega_i)\} - \Im\left\{\text{E}[Z(\omega_i|\mathbf{z})]\right\}\right)^2}.\end{aligned} \tag{7}$$

It is important to note that equation (7) is not used during the optimisation process but serves to provide feedback on the fit's accuracy.

In the plots' analysis, the pole number estimation remains stable. For intervals with high fuel utilisation, the estimated number of poles exhibits changes. These changes

do not suggest that a simpler model is valid in these intervals, as the root-mean-square error also increases. As indicated by [9], during these intervals, the Kramers-Kronig relations break down, signifying that the fuel cell is not operating within a linear regime, undermining the applicability of ECMs.

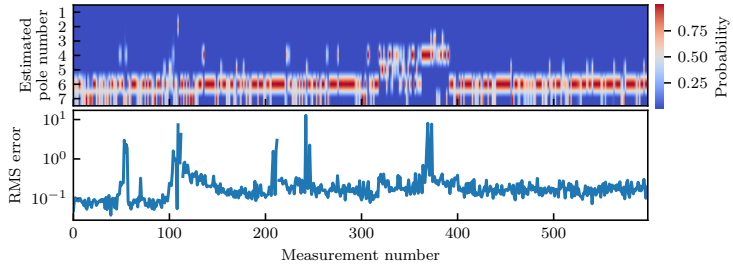


Figure 3: Pole evolution on experimental data

5 Conclusion

This paper introduces a statistical framework for analyzing EIS data, incorporating KL divergence, wavelet transforms, and VB inference alongside DRBS excitation. This combination allows faster and more reliable impedance measurements while providing detailed insights into electrochemical systems, such as lithium-ion batteries and SOFCs. The use of KL divergence enhances the detection of changes in impedance spectra, offering a sensitive tool for early identification of system degradation.

The proposed approach also uses VB inference for efficient model selection, accounting for the variability and uncertainty in the impedance data. Experimental validation on SOFCs and lithium-ion batteries demonstrates the accuracy and robustness of the framework across a wide range of operating conditions. These results highlight the framework's practical applicability for real-time diagnostics and monitoring of electrochemical systems.

References

[1] Gjorgji Nusev, Đani Juricic, Miran Gaberšček, Jože Moškon, and Pavle Boškoski. Fast impedance measurement of li-ion battery using discrete random binary excitation and wavelet transform. *IEEE Access*, 9:46152–46165, 2021.

[2] Pavle Boškoski, Andrej Debenjak, and Biljana Mileva Boshkoska. *Fast Electrochemical Impedance Spectroscopy*. Springer International Publishing, 2017.

[3] Jiapeng Liu and Francesco Ciucci. The gaussian process distribution of relaxation times: A machine learning tool for the analysis and prediction of electrochemical impedance spectroscopy data. *Electrochimica Acta*, 331:135316, January 2020.

[4] Vanja Subotić, Benjamin Königshofer, Đani Juričić, Mihails Kusnezoff, Hartmuth Schröttner, Christoph Hochenauer, and Pavle Boškoski. Detailed insight into pro-

cesses of reversible solid oxide cells and stacks using drt analysis. *Energy Conversion and Management*, 226:113509, December 2020.

[5] Pavle Boškoski, Luka Žnidarič, Žiga Gradišar, and Vanja Subotić. Probabilistic deconvolution for electrochemical impedance through variational bayesian inference. *Journal of Power Sources*, 622:235359, December 2024.

[6] Q. Fang, L. Blum, and N. H. Menzler. Performance and degradation of solid oxide electrolysis cells in stack. *Journal of The Electrochemical Society*, 162:F907–F912, 2015.

[7] A. Leonide, V. Sonn, A. Weber, and E. Ivers-Tiffée. Evaluation and modeling of the cell resistance in anode-supported solid oxide fuel cells. *Journal of The Electrochemical Society*, 155:B36, 2008.

[8] V. Sonn, A. Leonide, and E. Ivers-Tiffée. Combined deconvolution and cnls fitting approach applied on the impedance response of technical Ni-8YSZ cermet electrodes. *Journal of The Electrochemical Society*, 155:B675, 2008.

[9] Gjorgji Nusev, Bertrand Morel, Julie Mougin, Đani Juričić, and Pavle Boškoski. Condition monitoring of solid oxide fuel cells by fast electrochemical impedance spectroscopy: A case example of detecting deficiencies in fuel supply. *Journal of Power Sources*, 489:229491, March 2021.

Comparative Analysis of Energy Consumption Prediction Models: LSTM, Interval-Based, and Hybrid Models

Jelena Joksimović[†‡], **Jure Kos**[†], **Krištof Debeljak**[†]
[†]Rudolfovo Science and Technology Center
Podbreznik 15, 8000 Novo mesto, Slovenia
{jelena.joksimovic, jure.kos, kristof.debeljak}@rudolfovo.eu
[‡]corresponding author

Abstract. *This paper presents a comparative analysis of three predictive approaches for 1-day ahead energy consumption forecasting in two different types of buildings: a business office and a residential building. Energy consumption data, originally collected in 1-minute intervals, was aggregated to 15-minute intervals for model feeding. The first approach utilizes a Long Short-Term Memory (LSTM) network with 96 prediction steps, using a sequence length of 288 (equivalent to three days of historical data). The second approach employs an interval-based methodology, wherein 96 separate LSTM models are trained for each 15-minute interval with a sequence length of 10, resulting in a more granular prediction process. Lastly, we present a hybrid model that integrates the interval-based approach with daily behavior clustering to enhance the predictive capability. The performance of these methods is evaluated based on the Root Mean Square Error (RMSE). Results demonstrate that while the interval-based approach (2) outperforms the full-sequence LSTM (1), the Hybrid approach (3) achieves the highest accuracy in both building types, showcasing its potential for more reliable energy consumption forecasting.*

Keywords. energy consumption, energy management, recurrent neural networks, LSTMs, k-Means, deep learning

1 Introduction

Accurate energy consumption forecasting is critical for optimizing energy management systems in both commercial and residential buildings. With increasing emphasis on sustainability and energy efficiency, the ability to provide reliable day-ahead predictions becomes invaluable. The variability in energy consumption patterns between business offices and residential buildings makes it crucial to develop models tailored to specific building types, enabling more effective energy management solutions.
Recent experiments in Deep Neural Networks, particularly Long Short-Term Memory (LSTM) networks, have shown significant promise for energy-related tasks [5, 4, 1, 6, 8, 3]. LSTMs excel at capturing temporal dependencies in sequential data, which is particularly important for energy forecasting. However, traditional single-step LSTM models that attempt to predict long sequences, such as 96 steps ahead (covering an entire day

with 15-minute resolution), often suffer from accumulated errors. This has been well-documented in studies where LSTM models tend to lose accuracy over long prediction windows due to the compounding of errors across multiple time steps [7]. To overcome this, an interval-based approach has been proposed in this article, where smaller intervals are predicted independently, allowing for more stable and precise results.
In an interval-based LSTM approach, the forecasting task is broken down into separate models for each 15-minute interval of the day. Each model predicts only the next 15 minutes, effectively reducing the prediction horizon from 96 steps to 1, which helps prevent error accumulation over longer sequences. By focusing on short-term predictions and concatenating the results into a full-day forecast, this method can offer superior accuracy compared to traditional LSTM models [9]. However, while this approach improves accuracy, it does not fully account for the behavioral patterns of energy usage that can repeat daily in many buildings.
To further enhance prediction accuracy, hybrid models that combine interval-based LSTM forecasting with behavioral clustering have been introduced. In these hybrid models, daily energy consumption patterns are first clustered based on similarity, allowing for the identification of recurring behavioral patterns. This clustering-based method, inspired by recent work on temporal-behavior coalescing [7], has shown promise in improving the robustness and accuracy of predictions.
In this study, we compare three approaches to energy consumption forecasting for both business office and residential buildings. The first approach is a traditional LSTM with long prediction sequences (96 steps ahead), the second approach utilizes an interval-based LSTM model where each 15-minute interval is predicted separately, and the third approach incorporates a Hybrid model that combines interval-based LSTM predictions with clustering to account for recurring behavioral patterns. The primary objective is to determine which approach offers the best day-ahead forecasting accuracy, using Root mean square error (RMSE) as the main evaluation metric, for office and residential building types.

2 Methods

2.1 LSTMs for temporal forecasting

LSTM is a state-of-the-art variant of Recurrent Neural Networks (RNN) and is widely recognized for its effectiveness in time series data modeling. RNNs often struggle with long-term dependencies due to the vanishing gradient problem, while LSTMs overcome this limitation by introducing memory cells capable of retaining information for extended periods. Each LSTM memory cell consists of three gates: an input gate, an output gate, and a forget gate. These gates regulate the flow of information, allowing the network to selectively remember or forget data across multiple time steps [4, 1]. The gates control the interactions between neighboring memory cells and the memory cell itself, as illustrated in Figure 1. In this work, we compare two approaches:

1. **Full-Day Predictions:** The first approach involved training a single LSTM model to predict energy consumption for the entire day. This model generates 96 predictions, where each prediction corresponds to a 15-minute interval, covering a full 24-hour period. The LSTM was trained on sequences of length 672, corresponding to the past 7 days, to capture long-term dependencies in energy consumption patterns. Although this approach allows for complete day-ahead forecasting in a

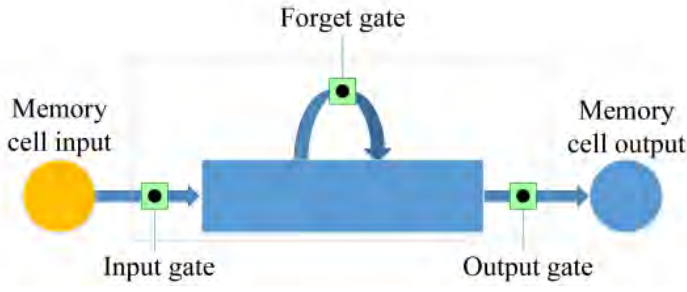


Figure 1: LSTM memory cell [2].

single pass, accumulated errors over long sequences diminished its accuracy, necessitating an alternative method. The results of this approach will be denoted as LSTM(1).

2. **Interval-Based Predictions:** To address the shortcomings of the full-day LSTM model, we adopted an interval-based approach. Instead of predicting the entire day in one go, we trained 96 separate LSTM models, each designed to predict one 15-minute interval (e.g., 00:00-00:15, 00:15-00:30, etc.), Each interval-based LSTM model used a time series of length 390 to predict the next 15-minute energy consumption. After generating 96 separate predictions, we concatenated them to form the complete day-ahead forecast. This method mitigates the error accumulation problem associated with long sequences. The results of this approach will be denoted as LSTM(2).

2.2 Hybrid Prediction Method: Daily consumption profile clustering + Temporal predictions

In addition to using LSTM(2) for temporal forecasting, we incorporated k-Means clustering to enhance prediction accuracy by accounting for recurring daily behavior patterns in energy consumption. We use k-Means clustering in this work, an unsupervised learning algorithm that groups similar data points into clusters based on their proximity in multi-dimensional space. For this study, we set k = 10, meaning daily energy consumption behaviors were grouped into 10 distinct clusters. Our chosen distance metric here is Euclidean distance. After calculating all 96 LSTM interval predictions and concatenating them into a single-day data vector, this vector was then processed through the cluster prediction algorithm, where the centroid of the corresponding cluster was identified. The centroid vector and the LSTM prediction vector were then averaged to generate the final hybrid predictions. Results of this approach will be denoted as Hybrid.

For model performance comparison we use root mean square error (RMSE), defined as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

where $\hat{y}_i$ is the predicted value of energy consumption, y_i is the actual value of energy consumption and n is the number of samples.

3 Results

The data used in this work consists of past energy consumption records (in Watts) for two buildings, C1 (office building) and C2 (residential building), over the course of 13 months with a 15-minute granularity. The key features used for model training are presented in Table 1. For LSTM(1) and LSTM(2), the optimal network parameters are obtained

Feature Type	**LSTM(1)**	**LSTM(2)**
Time-based features	week_day (1 to 7) is_weekend {0, 1} month (1 to 12) day (1 to 365) hour (0 to 23)	week_day (1 to 7) is_weekend {0, 1} month (1 to 12)
Past consumption	lags (1 to 4)	lags (1 to 30) (past days)
Sine/Cosine transformation	hour transformations (-1, 1)	-
Consumption timing flags	is_peak_hour {0, 1} is_low_usage_hour {0, 1}	-
Rolling mean	rolling mean per 1h rolling mean per 2h	-
Season	-	season (0 to 3)

Table 1: Set of features for models (1) and (2).

via neural architecture search, and can be found in Table 2. For the Hybrid approach,

Parameters	**Model**	
	LSTM (1)	LSTM (2)
Time series length	37,440	390
Sequence length	288	10
Prediction timestep	96	1
Number of layers	1	2
Number of units	100	150
Learning rate	0.001	0.001
Activation function	ReLU	ReLU
Loss function	mean squared error	mean squared error
Batch size	32	32
Epochs	20	20

Table 2: The optimal parameter values of the LSTM models. Interval approach (LSTM (2)) introduced opportunity for 2 LSTM layers, because with a decrease of prediction step from 96 to 1 our network was a way less computationally demanding.

we employed k-Means clustering with 10 clusters to identify days with similar energy usage profiles. The data was standardized using StandardScaler. These clusters were used to improve predictive accuracy by incorporating behavioral patterns into the temporal predictions. In Figure 2, we show the 10 daily load profiles of an example building [1].

[1]These clusters are illustrative examples and do not reflect real client data. Data can be provided upon request. Clusters in the figure belong to the Labtop energy consumption - the technical laboratory in Rudolfovo

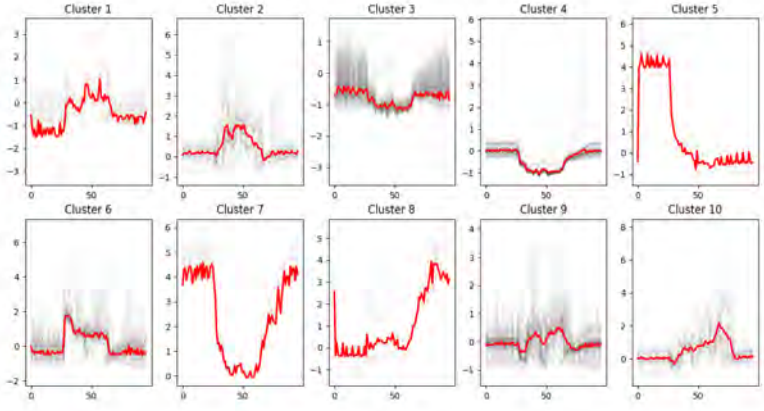


Figure 2: 10 clusters of daily energy load profiles example. Centroids are represented by red color, while all the actual time series are in grey.

The error metrics in kWh are presented in Table 3. The comparison between LSTM(1),

RMSE	**LSTM(1)**	**LSTM(2)**	**Hybrid**
C1	1.13	0.56	0.45
C2	0.46	0.39	0.26

Table 3: Evaluation metrics of LSTM full day predictions approach (1), LSTM interval predictions (2), and Hybrid predictions which average (2) with the daily consumption profile cluster centroid.

LSTM(2), and the Hybrid approach shows clear improvements in prediction accuracy with the interval-based and hybrid methods. LSTM(1), which uses a single LSTM model to predict all 96 steps for day-ahead consumption, struggled with accumulated errors and limited model capacity, leading to higher RMSE values (1.13 kWh for C1 and 0.46 kWh for C2). This was due to the complexity of predicting long sequences while minimizing the error on the daily level, while it it important to minimize the error on the 15-minutes level. LSTM(2), the interval-based approach, broke the prediction task into 96 individual one-step forecasts where the LSTM was trained separately for every interval, enabling the use of a two-layer LSTM architecture. This improved both accuracy and computational efficiency, reducing RMSE to 0.56 kWh for C1 and 0.39 kWh for C2. By addressing the issue of error accumulation over long sequences, LSTM(2) outperformed LSTM(1). The Hybrid method, which averaged LSTM(2) predictions with the centroids of daily behavior clusters, yielded the best performance, with RMSE values of 0.45 kWh for C1 and 0.26 kWh for C2. Training times for LSTM(2) and the Hybrid approach were approximately 2 hours and 20 minutes, with models trained only once per month, while generating predictions takes just under one minute. Although training LSTM(1) took 1 hour and 28 minutes, the decrease in prediction error achieved by LSTM(2) and the Hybrid approach justifies the additional training time. Some daily examples of predicted vs. actual values

are presented in the Figure 3. [2]

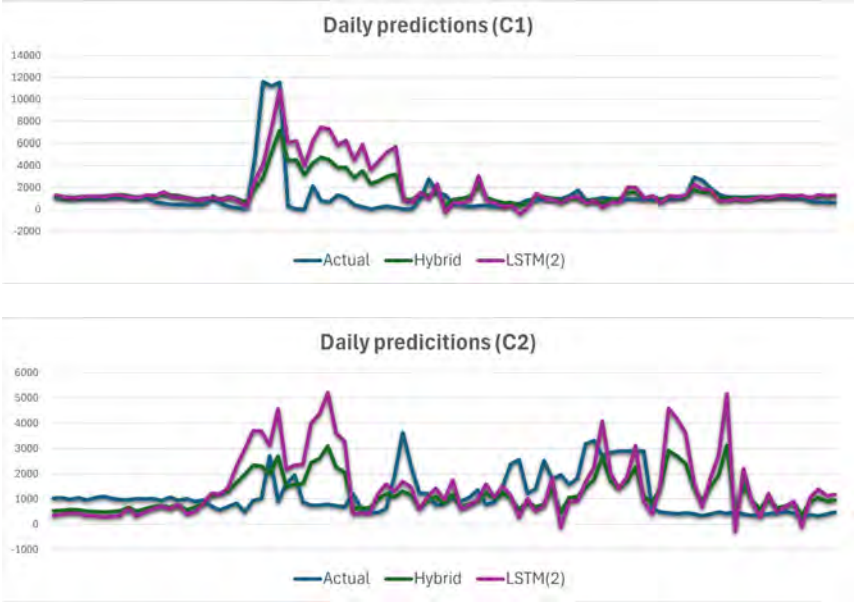


Figure 3: Predictions on the serving dataset (unseen data) in Watts.

4 Discussion and future work

Day-ahead forecasting, particularly when predicting energy consumption over a full 24-hour period, poses a significant challenge for LSTM models due to the need for 96 consecutive prediction steps (each representing 15-minute intervals). As the sequence length increases, maintaining both accuracy and computational optimum becomes increasingly difficult. The comparison between LSTM(1), LSTM(2), and the Hybrid approach demonstrates notable improvements in prediction accuracy with the interval-based and hybrid methods. The integration of behavioral patterns has proven to be useful in enhancing the accuracy of the predictions. Future work will focus on incorporating weather features and distinguishing between different energy load sources within the building. Currently, we have only worked with the building's total consumption, without breaking down individual contributors. Additionally, future steps will involve implementing Bidirectional LSTMs (BiLSTMs) [5] and utilizing HPC resources to explore deeper architectures, with the aim of further improving prediction accuracy.

References

[1] Almalaq, A., Zhang, J.J.: Evolutionary deep learning-based energy consumption prediction for buildings. IEEE Access **7**, 1520–1531 (2019). https://doi.org/10.1109/ACCESS.2018.2887023

[2] We were limited in the amount of data for sharing, for the given buildings.

[2] Bao, W., Yue, J., Rao, Y.: A deep learning framework for financial time series using stacked autoencoders and long-short term memory. PLoS ONE **12** (07 2017). https://doi.org/10.1371/journal.pone.0180944

[3] Cegovnik, T., Dobrovoljc, A., Povh, J., Rogar, M., Tomsic, P.: Electricity consumption prediction using artificial intelligence. Central European Journal of Operations Research **31** (05 2023). https://doi.org/10.1007/s10100-023-00844-6

[4] Hochreiter, S., Schmidhuber, J.: Long Short-Term Memory. Neural Computation **9**(8), 1735–1780 (11 1997). https://doi.org/10.1162/neco.1997.9.8.1735, https://doi.org/10.1162/neco.1997.9.8.1735

[5] Joksimović, J., Kos, J., Debeljak, K.s., Povh, J.: Recurrent neural networks for energy management systems: A case study. In: The Third Serbian International Conference on Applied Artificial Intelligence (SICAAI) (05 2024), subimtted.

[6] Kim, J., Moon, J., Hwang, E., Kang, P.: Recurrent inception convolution neural network for multi short-term load forecasting. Energy and Buildings **194**, 328–341 (2019). https://doi.org/https://doi.org/10.1016/j.enbuild.2019.04.034, https://www.sciencedirect.com/science/article/pii/S0378778819308072

[7] Liu, C., Li, F., Zhang, C., Sun, B., Zhang, G.: A day-ahead prediction method for high-resolution electricity consumption in residential units. Energy **265**, 125999 (2023). https://doi.org/https://doi.org/10.1016/j.energy.2022.125999, https://www.sciencedirect.com/science/article/pii/S0360544222028857

[8] Lu, Y., Sheng, B., Fu, G., Luo, R., Chen, G., Huang, Y.: Prophet-eemd-lstm based method for predicting energy consumption in the paint workshop. Applied Soft Computing **143**, 110447 (2023). https://doi.org/https://doi.org/10.1016/j.asoc.2023.110447, https://www.sciencedirect.com/science/article/pii/S1568494623004659

[9] Wang, J.Q., Du, Y., Wang, J.: LSTM based long-term energy consumption prediction with periodicity. Energy **197**(C) (2020). https://doi.org/10.1016/j.energy.2020.117, https://ideas.repec.org/a/eee/energy/v197y2020ics0360544220303042.html

Knowledge Transfer for Improved Detection of Electrical Insulators in Power Grid Photographs: Adapting Pre-trained Models to Slovenian Infrastructure

Andrej Trunkl,
GDI Ljubljana d.o.o.
Šmartinska cesta 106, 1000 Ljubljana

Pavle Boškoski
Faculty of Information Studies
Ljubljanska cesta 31A, 8000 Novo mesto, Slovenia,
Jožef Stefan Institute
Jamova cesta 39, Ljubljana, Slovenia

{andrej.trunkl@gdi.net, pavle.boskoski@fis.unm.si}

Abstract: *A country's power grid is very complex, and its operation must be continuous, with as few interruptions as possible. To maintain uninterrupted service, regular inspections of network assets are necessary. Due to the time-consuming nature of these inspections, drones are used to photograph network assets. The volume of photographs taken is substantial, making manual review and spatial placement of these images a significant challenge. The aim of the research is to enhance the efficiency of monitoring network conditions by developing methods for more accurate and faster detection and analysis of infrastructure, thereby improving the reliability and safety of the grid.*

Key Words*: spatial placement, infrastructure detection, photo pre-processing, pre-trained models, knowledge transfer*

1 Introduction

A power grid is a complex network that comprises a multitude of components that are subject to regular recording and monitoring. With many power poles and multiple recordings from each one, the number of stored photographs increases rapidly. Advances in data capture technologies and improvements in sensor accuracy on drones allow the company that manages the network to effectively monitor assets on the ground from an aerial perspective. However, the large volume of photographs poses a challenge in manually reviewing and identifying relevant image features. To address this problem, we can turn to artificial intelligence and machine learning technologies, particularly neural networks. Developing large-scale neural networks is a complex and costly process that requires a wealth of data for effective learning. Training models to label image material

is time/consuming and expensive. Off-the-shelf models, such as Single Shot Detector (SSD), RetinaNet, YOLO[1], and many others can be found on the market. However, the performance of such pre-built machine learning models may vary depending on the geographical area and the specific characteristics of objects, such as insulators in use in the Republic of Slovenia. For this reason, the knowledge transfer techniques from a pre-trained model are often used. This is a technique in machine learning where a model that has already been trained on one task is reused for another closely related task [2]. This process allows for faster and more efficient training, as the model already understands the objective of discovery, and this knowledge can be transferred to the new task.
In what follows, we will evaluate the performance of the pre-trained models developed by ESRI for the detection of electrical insulators in photographs of the power grid assets in use in Republic of Slovenia. The models are based on publicly available data published in the ESRI ArcGIS Living Atlas of the World [3]

The selected ESRI models will then be adapted to the electrical insulator-specificities in ELES photographs. The adaptation will be done using a smaller dataset than was needed to train the original ESRI model, which is the essence of knowledge transfer.

2 Basic properties of the models

Object detection in Ai and machine learning introduce the term bounding box. This is the minimal rectangle encompassing the object to detect. The bounding boxes are very convenient for training the models and for inference visualization. The prevalence of bounding boxes can be partially attributed to common metrics for object detection [4] and ease of computation. The down side of this is that the bounding box gives only a coarse location of the object and detection can be heavily influenced by background content or uninformative foreground areas as seen on the right side of Figure 1. The RepPoints approach is different to the bounding box approach. In the RepPoints it learn to adaptively position a set of points over an objects spatial extent in a meaner that circumscribes it as seen on the left side of Figure 1. [4]
The RepPoints are learned in a top-down fashion from the input image/object features, allowing for end-to-end training and producing fine-grained localization without additional supervision [4]

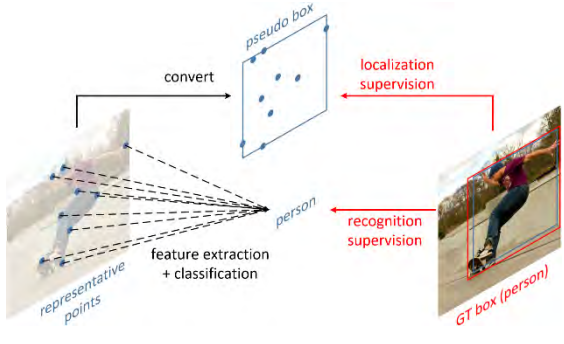


Figure 1: RepPoints operational representation [4]

RepPoints learning of representation points position is driven by two loss functions, the localization loss and the recognition loss. For localization loss RepPoints converts the representation points into pseudo bounding box and compares it with the ground truth bounding box. When the training is driven by both loss functions the extreme and semantical points of objects are automatically learned [4].

2.2 Model selection and structure.

For this research, two pre-trained models from ESRi Inc. were employed. Both models are based on the RepPoints architecture, implemented within the mmdetection framework. The mmdetection toolbox is built upon the PyTorch deep learning platform. Models in this toolbox are comprised of several key components: had, backbone, neck, region of interest (RoI) extractor, and loss.

The Head component is responsible for executing specific tasks, such as feature predictions. The Backbone typically consists of a fully connected network designed to extract feature maps. The neck acts as a bridge between the backbone and head. The RoI extractor identifies and extracts relevant regions of interest from the feature map. Finally, the loss component calculates the discrepancy between predicted and ground-truth values.

The two fundamental pre-trained models utilized by ESRi are characterized by the following components: a RepPointsHead head, a ResNeXt backbone, an FPN neck, and a SmoothL1Loss function for both head losses.

This model does not implement a feature extractor since the RepPointsHead model already extracts features.

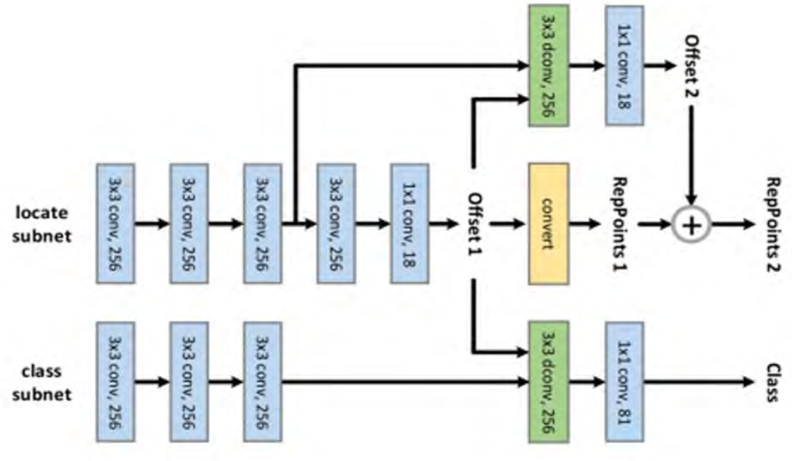


Figure 2: model's head part [4]

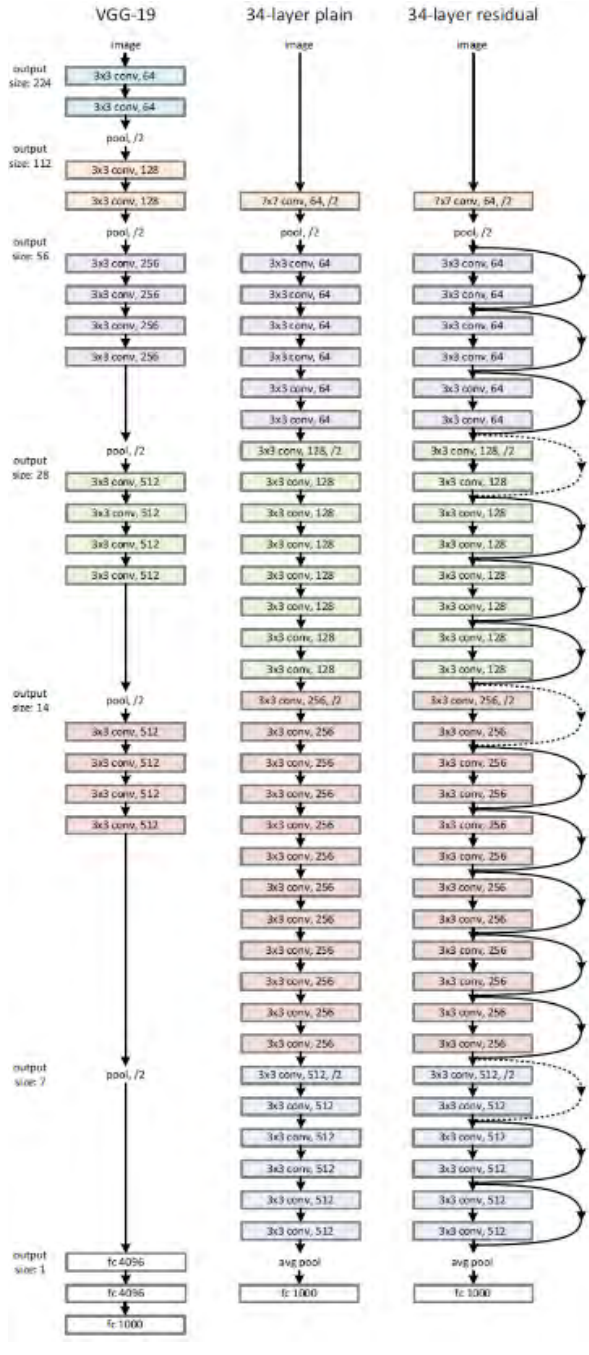


Figure 3: model's backbone [5]

The model uses the Momentum (0.875) optimizer and a learning rate (LR) of 0.256 for the batch size 256; if the size of the batch is smaller the LR gets scaled down [6].

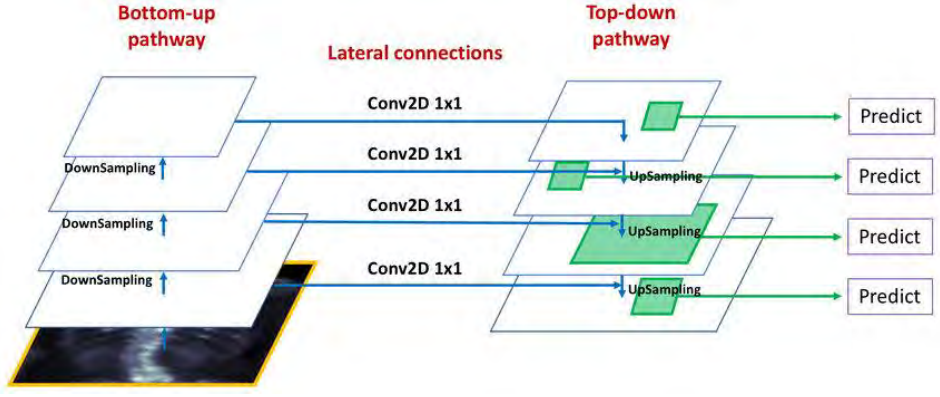


Figure 4: model's neck [7]

The FPN is implemented as described in Feature Pyramid Networks for Object Detection [8]

2.2 IDD Insulator defect detection

ESRI's Insulator Defect Detection (IDD) model is a pre-trained deep learning model designed to identify insulators and categorize defects within high-resolution directed images of insulator arrays mounted on transmission towers. A training dataset comprising 1596 photographs of porcelain insulators and a test set of 88 images were utilized for model development.

The IDD model classified objects into four distinct categories: 1) broken, 2) flashover damage, 3) string, and 4) no issue. Categories 1, 2, and 4 pertain to specific insulator components, while category 3 refers to the entire insulator string.

2.3 H-Structure detection model

The model is designed to detect H-structures and their constituent components within high-resolution directed images. This model can automate the identification of H-frame structures.

A dataset of 11,483 photographs of H-shaped columns was used for training, with 8,038 images allocated for training and 3,445 for testing. Objects were categorized into five classes: 1) H-structure, 2) pole, 3) crossarm, 4) insulator, and 5) X-brace. Only class 4 pertains to insulators; the remaining classes refer to the column members.

4 Transfer learning

Transfer learning aims to improve the current task by relating it to other tasks performed at different periods but from a related source domain. Transfer learning is most applicable when there is limited amount of source data available [2].

As described in the previous section, both models are trained for different categories of energy assets or different types of insulators (ceramic insulators). The Slovenian insulators differ from those on which the ESRI models were trained in terms of shape, size or material, which may affect the detection accuracy. Adaptation of the models to the local environment is crucial to improve their accuracy and reliability.

The aim of the knowledge transfer on these models was to create models that can identify from photographs the insulators used in the Slovenian high voltage electricity network. In Republic of Slovenia, glass and composite insulators are currently the most used insulators in high voltage grid, and the models were trained on porcelain insulators.

5 Data description

The entity responsible for operating the Slovenian portable power grid provided 240 photographs of electrical pylons equipped with insulators. These insulators are classified into two types: composite and glass.

The open-source labeling tool, Label Studio, was employed to annotate the objects within these images [10]. During the labeling process, two distinct categories were utilized to differentiate insulators based on their material: glass insulator and composite insulator. The object annotations were subsequently exported in the Pascal VOC format, a widely adopted standard for object annotation in computer vision model training [11].

Owing to the disparity in class labels between the exported annotations and the models, modifications were necessary. These modifications included the inclusion of additional, non-present class labels and the renaming and grouping of existing classes. This restructuring ensured that both glass and composite insulators were unified into a single "string" class for the IDD model and an "insulator" class for the H model.

7 Model parameters tuning and metrics

7.1 Model parameter tuning

The training process required the adjustment of two hyperparameters: network freezing and learning rate.

To investigate the impact of these hyperparameters, three distinct training configurations were employed:

- Exclusive unfreezing of the initial portion of the model (the head), utilizing the default freeze function.
- Freezing of the lower 30% of the model layers.
- Complete unfreezing of the entire model.

Following the preparation of the model or a portion thereof for training (fine-tuning), the optimal learning rate was determined using the methodology outlined by (L. N. Smith [9]).

This methodology, known as cyclical learning rate (CLR), involves the specification of minimum and maximum learning rate limits, as well as a step size. The step size represents the number of iterations (or epochs) utilized for each step. A cycle comprises two such steps: one where the learning rate increases linearly from the minimum to the maximum, and another where it decreases linearly. Smith [10]investigated various methods for varying the learning rate between these thresholds, concluding that they were equivalent and recommending the simplest approach: linear variation of the learning rate. A similar finding was reported by (L. N. Smith [9]), who suggested discrete jumps.

7.2 Used metrics

The primary metric employed for testing and evaluating the model's performance is average precision, which calculates the precision value at various recall values ranging from 0 to 1 [11].

Precision quantifies the accuracy of the model's predictions, while recall measures the model's effectiveness in identifying positive (true) instances. The Intersection over Union (IoU) metric calculates the Jaccard index between the ground truth (annotated) bounding box and the predicted bounding box. The IoU threshold, representing the minimum required overlap between the bounding boxes for metric computation, is typically set to 0.5, although there is no universally accepted value.

The average precision is the area under the precision-recall curve. In the case of object detection, the notions of recall and precision are calculated based on the intersection over the union as shown in Figure 5 [11].

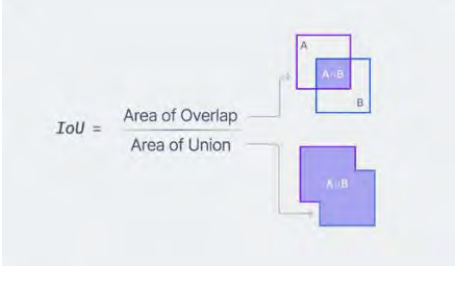


Figure 5: Intersect over union [12]

There are two additional concepts that have to be considered: the ground truth boxes and the detected boxes. Generally, the latter is a subset of the former. Taking all this into consideration, true positive values are those for which IoU values are greater than a certain threshold. Precision then becomes the ratio of the true positives and all detected

boxes. The recall is the ratio of true positives and all ground truth boxes.

7 Results

Initial validation was conducted using the pre-trained ESRI models. Fifteen percent of the entire training dataset was randomly selected and used for insulator detection. After the detection process, the predictive power of the model was evaluated using the Average Precision [11] metric with the IoU threshold 0.2 (the IoU threshold is set the same as in the model evaluation). The results for the original models without additional training are presented in the table below.

Model	Average Precision Score for class 3 and class 4
IDD	0.2574
H	0.3697

Table 1: Insulator detection results using the IOU metric for class 3 (IDD model) and class 4 (H model)

Knowledge Transfer Models

Using various approaches of knowledge transfer for the two models, IDD and H, different strategies for "freezing" and "learning rate" were applied to achieve the results shown in Table 1.

- **Freezing**: This refers to whether the weights in the model were frozen during training. "Freeze" means the weights were completely frozen, "Freeze_to(82) – 30%" means the weights were frozen up to a certain layer (in this case, up to the 82nd layer, representing 30% of the model), and "Unfreeze" means all weights were unfrozen and could change during training.
- **Learning Rate**: This is the speed at which the model learns. The smaller the value, the slower the learning process.

Average Precision Score (APS): This metric evaluates the average precision of the model in classification tasks, based on the IoU metric.

Model Name	Freezing	Learning Rate	Average Precision Score for classes 3 and 4
IDD	freeze	0.000091	**1.0**
IDD	freeze_to(82) – 30%	0.004365	**0.818181**
IDD	unfreeze	0.0006927	**0.909091**
H	freeze	0.000575	**0.900826**
H	freeze_to(82) – 30%	0.003020	**0.727273**
H	unfreeze	0.0006918	0.727273

Table 2: Model training parameters and achieved precision

Validation of the IDD Default Freeze Model

For the selected IDD default freeze model, cross-validation was performed with fifteen iterations. The data was divided as follows:
15% of the images from the primary dataset were selected and stored as the test set. The remaining data was used for training the model.
The test results for each step of the cross-validation are presented in Table 4.

Model	Average Precision Score for class 3	Variance
IDD	0.6354	**0.0077**

Table 3: Overall Accuracy Score for the IDD Model

The metric of precision and recall were calculated for all fifteen iterations for the model IDD. The results of the calculation of these metrics are listed in Table 4.

Model IDD	Mean value	Variance
Precision	0.918	0.0007
Recall	0.661	0.0077

Table 4: Cross validation - precision and recall

The results of the pre-trained ESRI models were less favorable, as they were trained on data that differed in shape, size, and material from those used by the company ELES, which affected detection accuracy. With knowledge transfer, the accuracy of these models improved.
From the results of the models based on knowledge transfer, it is evident that the IDD default freeze and H default freeze models performed the best. The IDD default freeze model achieved an accuracy score of 1.0 for category 3 (string).

The IDD model, with frozen weights and a very low learning rate, achieved perfect accuracy. When the weights were partially or fully unfrozen, the accuracy decreased but remained high.

For the H model, the results showed that freezing the weights or partially unfreezing them did not significantly affect accuracy, while fully unfreezing the weights led to the same accuracy as partial unfreezing.

Models with a lower learning rate and 'Freeze' setting generally achieved higher accuracy, indicating that it might be better to maintain greater stability during training for these models, rather than allowing larger weight adjustments during the learning process. Perfect accuracy of 1.0 for the IDD model with 'Freeze' and a very low learning rate might be suspicious, as achieving perfect accuracy is very difficult in practice, especially on complex or real-world datasets. This could indicate some issues, such as:

- **Overfitting**: The model might be too closely fitted to the training data and may not generalize well to new, unseen data.

- **Data Error**: There could be issues with the dataset, such as labelling errors or inadequate pre-processing.

The data was checked and no errors were found. Upon reviewing the learning and validation metrics, it was found that the loss function curves fit well, and the shape of both curves did not indicate overfitting. Therefore, the model was still marked as usable.

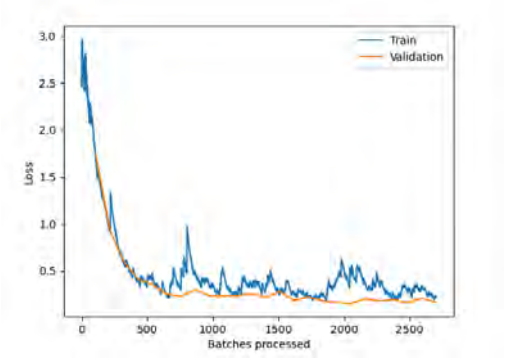


Figure 6: IDD, Default Freeze

Since it was determined that the IDD model performs better, a cross-validation was conducted to evaluate it performance. The results confirmed that the model performs well.

Figure 7:Predictions of the IDD Default Freeze Model

The IDD default freeze model achieved slightly lower accuracy in the cross-validation compared to the predicted accuracy from model validation. The model's output are multiple bounding boxes of the same class that are within or encompass the ground truth bounding box. There for, the same object is detected multiple times. To test this, a procedure was developed that searched overlapping bounding boxes and combined them in a single bounding box. With this procedure the calculated average precision increased greatly.

After selecting the model, cross-validation was performed as described in the section on validating the IDD default freeze model. It was found that the model had improved and is sufficiently good for detecting insulators.

More accurate results would require a larger number of labeled photographs with different backgrounds, and the development of a bounding box merging logic to prevent one insulator from being detected multiple times.

8 Conclusion

This study explored the application of knowledge transfer techniques to adapt pre-trained models for insulator and H-structure detection in the Slovenian power grid infrastructure. By fine-tuning ESRI models, originally developed for different geographical areas and asset types, we demonstrated that the adapted models are capable of accurately detecting insulators specific to the Slovenian context, including composite and glass insulators, which differ from the porcelain insulators used in the original training data.

The performance evaluation using the Intersection over Union (IoU) metric showed that while the initial pre-trained models struggled with the detection accuracy, the application of transfer learning significantly improved the results. The best-performing model, IDD with a freeze setting and low learning rate, achieved near-perfect detection accuracy, suggesting a strong fit for the task. However, this result warrants caution due to the potential risk of overfitting, which was mitigated through cross-validation and detailed review of the learning metrics.

Despite the promising results, further improvement is necessary to enhance generalization. This could be achieved by increasing the dataset size and diversity, especially with more varied backgrounds, and by refining the post-processing methods to address the issue of redundant bounding boxes. Additionally, future work should focus on optimizing the balance between precision and recall, and on fine-tuning hyperparameters to improve the robustness of the models when applied to new and unseen data.

In conclusion, the transfer learning approach has proven effective in adapting the ESRI models to local conditions, providing a valuable tool for automating the monitoring of Slovenia's high-voltage infrastructure. With additional data and further model enhancements, this method can become an integral part of asset management practices for power grids, contributing to improved efficiency and reliability.

9 References

[1] "ArcGIS API for Python," ArcGIS API for Python. Accessed: Aug. 01, 2024. [Online]. Available: https://developers.arcgis.com/python/

[2] A. Hosna, E. Merry, J. Gyalmo, Z. Alom, Z. Aung, and M. A. Azim, "Transfer learning: a friendly introduction," *J Big Data*, vol. 9, no. 1, Art. no. 1, Dec. 2022, doi: 10.1186/s40537-022-00652-w.

[3] A. L. A. of the World, "ArcGIS Living Atlas of the World." Accessed: Jul. 23, 2024. [Online]. Available: https://livingatlas.arcgis.com/

[4] Z. Yang, S. Liu, H. Hu, L. Wang, and S. Lin, "RepPoints: Point Set Representation for Object Detection," arXiv.org. Accessed: Jul. 23, 2024. [Online]. Available: https://arxiv.org/abs/1904.11490v2

[5] A. OTHMAN and C. Direkoglu, "Evrişimli Sinir Ağlarını Kullanarak Müzik Notasyonunu Tanıma," *European Journal of Science and Technology*, Nov. 2020, doi: 10.31590/ejosat.823266.

[6] “DeepLearningExamples/PyTorch/Classification/ConvNets/resnext101-32x4d at master · NVIDIA/DeepLearningExamples,” GitHub. Accessed: Oct. 07, 2024. [Online]. Available: https://github.com/NVIDIA/DeepLearningExamples/tree/master/PyTorch/Classification/ConvNets/resnext101-32x4d

[7] T. Apiparakoon *et al.*, “MaligNet: Semisupervised Learning for Bone Lesion Instance Segmentation Using Bone Scintigraphy,” *IEEE Access*, vol. PP, pp. 1–1, Feb. 2020, doi: 10.1109/ACCESS.2020.2971391.

[8] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, “Feature Pyramid Networks for Object Detection,” arXiv.org. Accessed: Oct. 07, 2024. [Online]. Available: https://arxiv.org/abs/1612.03144v2

[9] L. N. Smith, “A disciplined approach to neural network hyper-parameters: Part 1 -- learning rate, batch size, momentum, and weight decay,” arXiv.org. Accessed: Aug. 02, 2024. [Online]. Available: https://arxiv.org/abs/1803.09820v2

[10] L. N. Smith, “Cyclical Learning Rates for Training Neural Networks,” Apr. 04, 2017, *arXiv*: arXiv:1506.01186. doi: 10.48550/arXiv.1506.01186.

[11] J. Hui, “mAP (mean Average Precision) for Object Detection,” Medium. Accessed: Oct. 07, 2024. [Online]. Available: https://jonathan-hui.medium.com/map-mean-average-precision-for-object-detection-45c121a31173

[12] “Intersection over Union (IoU): Definition, Calculation, Code.” Accessed: Sep. 04, 2024. [Online]. Available: https://www.v7labs.com/blog/intersection-over-union-guide, https://www.v7labs.com/blog/intersection-over-union-guide

Parameters devise for modelling real-time monitoring of magnetic field for the purpose of a smart city using a mobile phone

Valerij Grasic
Telekom Slovenia
Cigaletova 17, 1000 Ljubljana, Slovenia
grasic.se@gmail.com

Biljana Mileva Boshkoska
Faculty of Information Studies
Ljubljanska cesta 31A, 8000 Novo mesto, Slovenia
Jožef Stefan Institute
Jamova cesta 39, 1000 Ljubljana
biljana.mileva@fis.unm.si

Abstract: *The transition to a fully digitized society requires the digitization of all segments. In the case of a smart environment, consideration of the magnetic field is gaining importance, especially in light of the recent developments in space weather and aurora phenomena. We present an initial study on the design of parameters corresponding to the layout of a model for monitoring the magnetic field in real time for smart cities, using a mobile phone, where the magnetic field includes all influences at a specific measurement location. The design of the model parameters is based on a detailed analysis of six different magnetic field datasets, collected from distinct environments: a static location, walking measurements, and measurements taken while driving a car, with both rural and urban settings considered for each scenario.*

Key Words*: Smart City, Smart Environment, Geomagnetic, Artificial Intelligence (AI), Magnetometer, CrowdMag*

1 Introduction

Electromagnetic fields, although invisible, are omnipresent. They are typically static but can change due to various elements in the surrounding area or events related to space weather. Recently, in May 2024, a significant solar event occurred when Earth was hit by a series of powerful Coronal Mass Ejections (CMEs) resulting from a large, magnetically complex sunspot cluster on the sun. These CMEs triggered G5-class geomagnetic storm. One positive outcome of this storm was the visibility of stunning auroras in regions where they are not normally seen, such as Slovenia. However, due to major changes in the Earth's magnetic field, the consequences were considerable, especially for technology-dependent systems, such as the operation of GPS systems [1].

Our goal is to devise a magnetic field measurement concept that a smart city user can utilize in real time with their mobile phone. Smart cities rely on the use and processing of data related to IoT (Internet of Things) devices. These sensor devices vary widely and include mobile phones and the applications running on them. They cover various aspects

of a smart city, such as smart mobility (e.g., parking management), smart infrastructure (e.g., speed measurement), and smart environment (e.g., weather, air quality). This enables both the display of device statuses and parameters and, eventually, the development of new digital city services based on this data. The segment of magnetic field monitoring is still in its infancy and forms a part of the smart environment. As for the magnetic field, various sources surround us: Earth has its own a magnetic field, as does the Sun, and any wire carrying electricity creates a magnetic field around it. The focus of this study is on the cumulative effect of all magnetic fields at a given location and time. The magnetic field at any specific location comprises different components and directions. In this study, we concentrate on the total magnetic field (T), along with its vertical (Z) and horizontal (H) components. We do not expect the magnetic field to have a completely fixed value, as it is already established in [1] that the magnetic field can change over time. However, we are interested in understanding the nature of these changes, their frequency and extent, as well as the underlying principles that apply.

In summary, our aim is to design a model for real-time magnetic monitoring that adapts to changing locations and situations. This work is the first step in this direction. First, we present the proposed parameters to enable magnetic field monitoring in smart cities; later, these parameters are evaluated on six sets of measurements - including static locations, walking, and driving scenarios in both rural and urban settings.

2 Problem description

2.1 Smart cities

By definition, a smart city is a place where traditional networks and services are made more efficient using digital and telecommunication technologies for the benefit of its residents and businesses, and we are still at the very beginning of smart city development [2]. Moreover, according to the UK government, the concept of smart cities is not static [3] because there is no absolute definition of a smart city or an endpoint. Instead, it is a process or a series of steps that make cities more alive and flexible, enabling them to respond more quickly to new challenges. The guidelines for smart cities are moving in the direction that problem-solving should be more innovative and intelligent [4], and that systems and solutions within them should be made more intelligent [5].

2.2 Sensing paradigm

An important question is how to increase situational awareness [6]. This question also involves intelligence, such as for incoming calls within a smart city [7, 8]. In the domain of crowdsensing, research has shifted from focusing solely on technology concepts to now emphasizing how they improve citizens' quality of life and utility [9]. Mobile Crowdsensing (MCS) has garnered considerable attention and emerged as a promising sensing paradigm [10], as it enables extensive coverage and contextual awareness, thereby providing valuable insights for various applications.

2.3 Magnetic field

Electromagnetic fields extend, on the one hand, to the Sun and into space (such as within our solar system) and, on the other hand, to the very core of the Earth. Several factors affect the magnetic field [11]. The geomagnetic field protects the Earth's atmosphere from charged particles in the solar wind [12]. Earth's magnetic field is complex and influenced by various factors [13]. These fields also change dynamically.

What occurs around Earth, as far as the magnetic field is concerned, also impacts what happens on Earth. The most important components are Earth's main magnetic field generated in the conducting, fluid outer core; the crustal field, generated in Earth's crust and upper mantle; and the combined disturbance field from electrical currents flowing in the upper atmosphere and magnetosphere, which induce electrical currents in the sea and ground [11]. Earth's main magnetic field dominates, accounting for over 95% of the field strength at Earth's surface. Secular variation is the slow change over time of the main magnetic field.

2.4 Space weather

The magnetosphere is the region of space surrounding Earth where the dominant magnetic field is Earth's own, rather than the magnetic field of interplanetary space. The magnetosphere is formed by the interaction of the solar wind with Earth's magnetic field. Conditions inside the magnetosphere are highly dynamic and create what we call "space weather" [14], which can affect technological systems and human activities. Important parameters for space weather include the Kp and Dst index [12].

Space weather influences atmospheric electricity [15]. The terrestrial ring current [16] is an electric current flowing toroidally around Earth, centred at the equatorial plane and at altitudes of 10,000–60,000 km. Changes in this current are responsible for global decreases in the Earth's surface magnetic field, known as geomagnetic storms. Intense geomagnetic storms can have severe effects on technological systems, such as disturbances or even permanent damage to telecommunication and navigation satellites, telecommunication cables, and power grids.

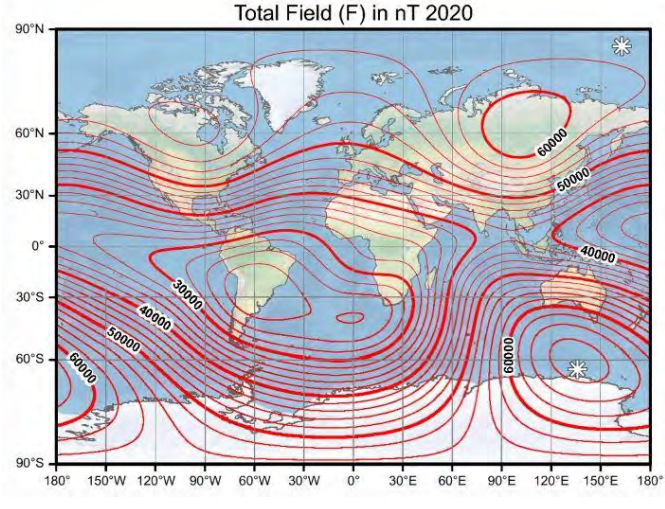


Figure 1: Magnetic Total Field (F) values for the whole Earth as defined by the IGRF model [18].

2.5 Measuring magnetic field

Rapid and long-term changes in the magnetic field are monitored by low-orbiting satellites and a global network of about 200 magnetic observatories [17]. Field models, computed from these measurements, provide the direction and strength of the magnetic

field at any desired location. Magnetometers can then be used to compute pointing directions, complementing the location information provided by the global positioning system (GPS). Figure 1 shows the value of the total magnetic field (F) for the whole Earth as defined by the IGRF (International Geomagnetic Reference Field) model [18]. According to estimates, the magnetic field on Earth ranges between 25 μT (at the equator) to 65 μT (at the poles).

Satellites measure the magnetic field around the Earth, while magnetometers at static locations monitor the field from the ground. Efforts are underway to enable real-time magnetic field monitoring for moving location [19]. This would allow for the development of magnetic models with much higher resolution, reaching a few meters rather than the approximately 3,000-kilometer resolution achievable with satellites alone. For Slovenia, using a theoretical model (as shown in Figure 1), the magnetic field is estimated to range between 40 and 60 μT.

Since 1848, real geomagnetic measurements have been conducted [20, 21, 22] in Slovenia, and more recently, measurements have also been taken by mobile phone [23]. Magnetometer data show that certain objects and infrastructure cause additional magnetic fields, whether from household devices with short-term, lower-intensity impact (e.g., iron, oven, hair dryer) or from surrounding infrastructure with a greater impact (e.g., railways, transformer stations, power lines, buildings). This influence can range from 0.1 to 0.2 μT or more, sometimes reaching up several μT. The vertical (Z) component has been shown to vary by a of +/-0.54 μT across Slovenia relative to the mean value [21]. According to [24], as of January 2011, the total magnetic field (T) in the western part of Slovenia (Ajdovščina) was measured at 47,59 μT.

2.6 Real-time magnetic field measurements for smart cities

The ability to measure magnetic fields in real time introduces a new paradigm in environmental monitoring. Historically, affordable magnetometers integrated into mobile devices were scarce. Even with the inclusion of magnetometers in phones, challenges persisted due to the complexity of Earth's magnetic field and the numerous factors that influence it [11, 13].

As society digitizes, real-time magnetic field measurements become essential, not only for tracking environmental conditions (as it is magnetic field) but also for developing new smart city services. These services should be intelligent [14, 15] and could support diverse applications, such as addressing GPS disruptions [1] or enhancing critical infrastructure protection [14]. Unlike traditional static measurements from satellites and stationary magnetometers, advancements in mobile technology now allow for short-range, high-resolution magnetic monitoring. This innovation, paired with smartphone-based IoT, enables users to capture magnetic data that is both location-specific and context-aware, providing dynamic tracking of magnetic field fluctuations.

Our research goes beyond measuring magnetic field strength to explore how these values shift over time and under varying conditions. We aim to understand the dynamics, boundaries, and underlying principles of these changes. Although magnetic fields might be expected to remain static at a given location, studies indicate that environmental factors can cause significant variations, sometimes reaching several microteslas (μT) [20-23].

Our objective is to incorporate these influences and assess their impact across specific scenarios, establishing a real-time monitoring profile that adapts to different locations and situations. This granularity facilitates anomaly detection and enables the creation of magnetic field maps, whether through crowd-sourced data (e.g., CrowdMag) or personalized mapping, thereby enriching smart city infrastructure.

3 Methodology description

3.1 Tools used

For magnetic measurements, an iPhone 13 equipped with a built-in magnetometer served as the primary tool. The magnetometer data was collected using the CrowdMag app (version 2.1.3) [19]. The measurements were then processed on a standard personal computer with an Intel i5 processor running Windows 11. Data from the app was saved in CSV format, emailed to the author, and later added to a repository [25]. Analysis was conducted using Python (version 3.12.4) with libraries such as NumPy, Pandas, Matplotlib, Scikit-learn, Plotly, and DateTime. The only limitation noted was the iPhone's standard deviation for magnetic measurements, which is 140 nT (0.14 μT) [26].

3.2 Measurements

To address the measurement approach, we aimed to capture data across several potential scenarios in a smart city, covering different locations and modes of movement while keeping the number of measurements manageable. This strategy provides broad insight into the magnetic field dynamics within the area of smart cities. Based on this plan, six sets of measurement were taken: two at static locations (S), two while walking (W), and two in a car (C), resulting in 28 measurements overall.

The first static location (S1-SE) was a rural house in the northern part of Slovenia, 1 m above ground on a flat surface; the second (S2-BK) was in the southern part of the country was on a hillside in nature. For walking, one location (W1-LJAP1) was in a capital city center, covering areas like bus and railway stations, buildings, underground garages, and different floor levels; the other (W2-SE) was rural (near S1-SE), passing power lines, a road, and forest. Car measurements included driving from the north (as from S1-SE) towards the capital (C1-SE-LJ), covering highways, rivers, power lines, and city buildings, while the second route (starting from S1-SE) ran along a rural, mountainous road eastward (C2-SE-DRETA), capturing the effects of varied topography.

No issues occurred during measurements. Sampling every 10 seconds proved optimal, and efforts were made to keep the phone consistently positioned throughout.

3.3 Magnetic components

CrowdMag provides three continuous geomagnetic field components [27], which were used for evaluation: H, Z, and F (representing the horizontal, vertical, and total magnetic field, respectively). F is a vector quantity (sum) derived from the H and V components of the magnetic field. Additionally, four other parameters are available [27] - D (declination), I (inclination), X (X component) and Y (Y component) – but these are only possible in the application for point measurement.

3.4 Analysis conducted

CrowdMag captures real-time measurements but only provides current values for each magnetic field component without any additional data processing. This processing was performed in Python, where basic statistical characteristics, including maximum and minimum values, average, and 95% intervals, were calculated. The main challenge was determining the optimal approach for evaluating these measurements.

4 Parameter devise proposal and associated parameter evaluation

Initially, a set of parameters is proposed to indicate changes in measurements that should be considered when monitoring the magnetic field in smart cities.

4.1 Parameters devise

The proposed parameters are organized into several segments (Seg), including General (G), Core (CO), Core Aggregated (CA), Environmental Context (EC), Space Weather Context (SC) and Model Output (M) data, as shown in Table 1.

Table 1: Parameters devise proposal for modeling magnetic fields in smart cities

Seg	Parameters, and their meaning
G	**Time, location, altitude, duration**: latitude, longitude, time spent in the area **Route**: speed, length, start/end time/location, altitude (start/end, average, minimum/maximum) **Name, repetitions:** name**,** number of repetitions (if multiple measurements) **Height**: altitude from the ground (due garages, undergrounds, height of floors)
CO	**T, Z, H:** actual values used as base measurements references **I, D, X, Y:** additional magnetic component values
CA	**T, Z, H avg:** average values over a short period (e.g., 10 to 60 seconds) to provide a stable magnetic representation **T, Z, H min/max:** indicate sudden changes in the field, allowing detection of anomalous spikes or drop**s** **T, Z, H 95% range:** capture typical magnetic variation by excluding outliers, providing a practical "normal range"
EC	**Nearby influences (NI) (up to 5 meters):** changes due to electronic devices, frequency of nearby sources, rocks, or walls **Farther influences (FI) (more than 5 meters):** effects from sources, like electric wires, transformers, trains, rivers or buildings **Locality:** contextual information such as city, countryside, settlement **Weather**: atmospheric events affecting EMF, e.g., lightning **Noise**: other influences, including movement or device orientation
SC	**Space weather:** solar flares, CME, geomagnetic storms (Kp, Dst), substorms
M	**Mode**: user's mode (static location (S), walking (W), or car (C)) **Alert**: anomaly alert, indicating unpredictable field behavior **Short-term anomalies (STA)**: peaks or sudden spikes **Long-term anomalies (LTA)**: gradual deviations from the rolling mean **State**: field state, such as normal, anomalous, increased, or decreased **Magnetic Map**: visual map showing the field over time

General parameters (G) cover measurement details such as time, location, altitude, duration, route, and height. The foundational parameters for the model, both Core (CO) and Core Aggregated (CA), consists of raw measurements (actual values are part of CO) and summarized statistics (average, min, max, 95% range values are part of CA) of magnetic components data (T, Z, H). The environmental context (EC) ensures that readings reflect the actual environment, rather than being skewed by known sources of

interference. For example, nearby influences, which may be short-lived, can be detected as sudden spikes or anomalies in the data using rolling averages or variance over short time intervals. Space weather parameters (SC) integrates solar storms and other magnetic disturbances (for explanation of parameters see [12, 14]), which can significantly affect readings.

The output data for the model (M) include the mode of movement (Mode), short or long anomalies, and possible alerts. Additionally, the output comprises the state of the magnetic field and the event map. The combination of core measurements (CO, CA), environmental (EC), and space weather context (SC) inputs can create a comprehensive geomagnetic profile for real-time monitoring.

4.2 Evaluation of the magnetic field

Using these parameters, six sets of measurements were obtained with an iPhone across three different environments: static location, walking, and car driving. These toy examples of magnetic field evaluations across diverse environments serve to demonstrate and validate the proposed parameter division.

4.3 Static location evaluation

The evaluation of static locations is shown in Figure 2 (within a household, labeled as S1-SE) and Figure 3 (on a slope in the countryside, labeled as S2-BK). Minor field fluctuations can be observed in Figure 2. In Figure 3, measurements taken in the countryside indicate greater fluctuations in the magnetic field, potentially due to the incline and the presence of large rocks in the area.

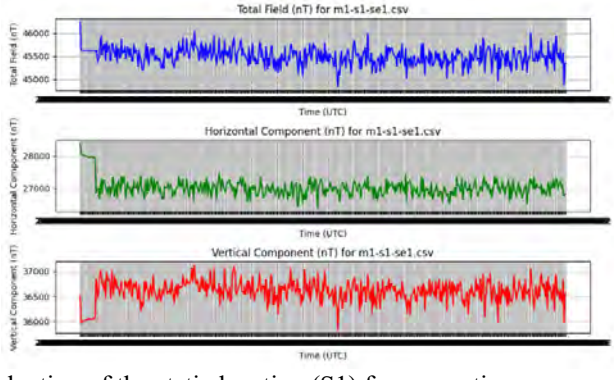


Figure 2: Evaluation of the static location (S1) for magnetic components T, H and V, conducted within a household setting.

4.4 Walking evaluation

Field evaluations while walking are presented in Figure 4 (urban environment within the city including buildings, underground garages and floors, W1-LJAP1) and Figure 5 (countryside with a settlement, W2-SE). In both cases, larger deviations are observed compared to static location measurements.

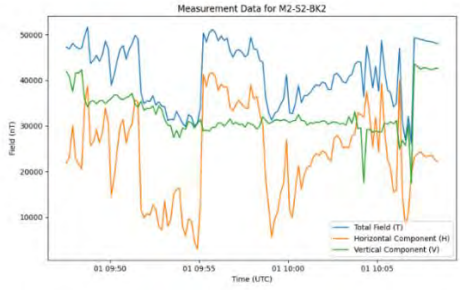


Figure 3: Evaluation of the static location (S2) for magnetic components T, H and V, taken in a mountainous countryside environment.

In Figure 4, the field deviations result from factors such as buildings, underground garages, and floors, where the fluctuations for each component (T, H, V) follow a similar pattern. Conversely, Figure 5 shows more stable measurements in the countryside. However, differences are noted between T and H, as they do not follow the same pattern, which highlights the necessity of considering all three components (T, H, and V) rather than just the total field (T).

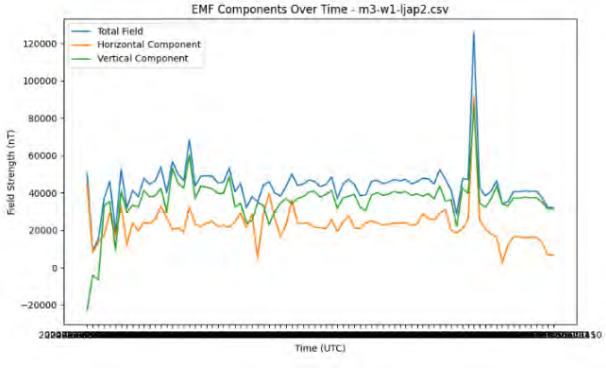


Figure 4: Evaluation of the magnetic field while walking (W1) for components T, H and V in an urban environment (within the city).

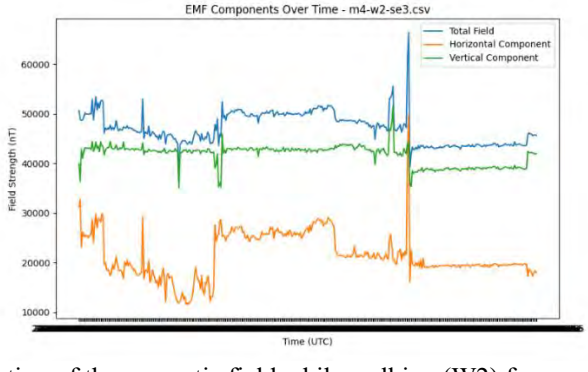


Figure 5: Evaluation of the magnetic field while walking (W2) for components T, H and V in a countryside setting with settlement.

Figure 6: Evaluation of the magnetic field while driving (C2) for components T, H and V in a mountainous area.

4.5 Car driving evaluation

The evaluation of magnetic field variations during car driving is shown in Figure 6 (driving over a mountain pass, C2-SE-DRETA) and Figure 7 (driving towards the capital city, C1-SE-LJ). These measurements exhibit the largest deviations compared to static locations or walking.

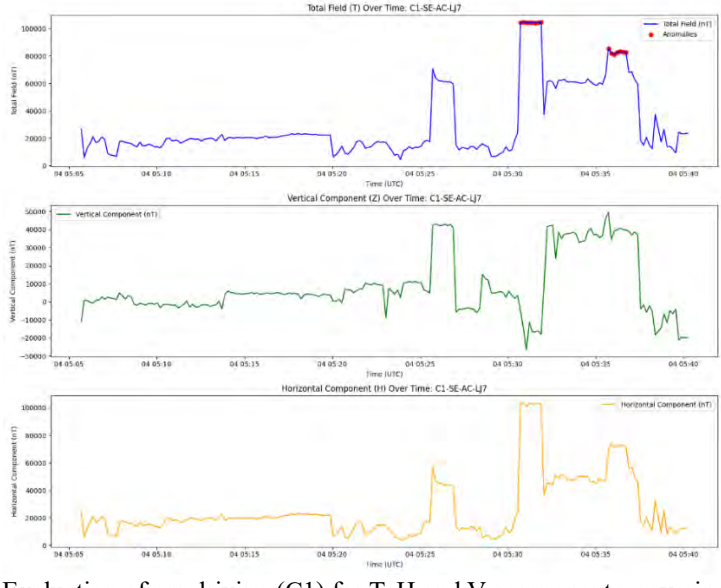


Figure 7: Evaluation of car driving (C1) for T, H and V components, covering the route from countryside to city.

However, the average value for T is significantly lower in these scenarios. Both figures also reveal anomalies where values exceed twice the standard deviation. The magnetic components (T, H, V) generally follow a similar pattern but with noticeable deviations. The influence of the car's structure, braking, and the engine is present in both cases. In Figure 6, additional effects from the hilly terrain (rocks, height) are evident. Figure 7 captures underpasses, tunnels, urban areas, and increased traffic density.

4.6 A model representation of the evaluated measurements using the proposed parameters

Table 2 presents a model representation of the evaluated measurements based on the proposed parameters. In this representation, only the total magnetic field (T) is utilized, with Core Aggregated (CA) values expressed in μT. Commas within the cells serve as separators between two sets of measurements.

Table 2: Evaluation of measurements using the proposed parameters

Seg	Param	Static	Walk	Car
G	Duration	30 to 90 minutes	30 to 90 minutes	30 to 90 minutes
	Name	S1-SE, S2-BK	W1-LJAP, W2-SE	C1-SE-LJ, C2-SE-DRETA
	Repetitions	7, 2	5, 4	7, 3
	Altitude	400 m, 430 m	400 m, 320 m	400/300 m, 400/1500/600 m
	Height	0	-10 m/+10 m, 0	0
CA	T avg	48475.1, 47251.4	42070.7, 47700.8	24469.6, 27009.2
	T min/max	42597.4/75499.6 26112.3/51641.7	9087.4/132462.9 36832.8/66421.0	1545.6/136474.2 4463.8/140886.9
	T 95%	74931.8, 49217.8	53387.4, 50232.5	58582.5, 48203.9
EC	NI	wall, TV set	buildings	car engine
	FI	house, rocks	substation, electric lines	mountains
	Locality	house, countryside	countryside, city	mountains, city, countryside
M	Mode	S	W	C
	Anomaly	no or rare	occasional	moderate
	STA	no	occasional	moderate
	LTA	rare	occasional (W1)	moderate
	Class	S1: regular S2: minor occasional anomalies	W1: some spikes W2: moderate deviations	moderate deviations

5 Conclusion

In this paper, we addressed the challenge of measuring geomagnetic fields, which are highly dynamic in nature, making this a timely and relevant research topic. By using a mobile phone magnetometer, we enabled real-time monitoring of these fields. The primary contribution of this study is the development of essential parameters for constructing a real-time magnetic field model, along with a tabular representation of data collected via a mobile phone magnetometer.

Based on our observations, we conclude that the proposed parameters are adequate and can be derived from the available measurements. Some parameters are defined independently, while others are calculated or based on observational manipulation. Automating the determination of these parameters in the future, and creating a database for environmental context values, would further clarify the actual field and support the development of a robust real-time magnetic field model for smart cities.

Our findings emphasize the potential of mobile phones in real-time magnetic field assessment and mapping, independent of satellite data or static-location magnetometers, presenting significant opportunities for future research and advancements in this field.

6 References

[1] Albeck-Ripka, L. Solar Storm Crashes GPS Systems Used by Some Farmers, Stalling Planting. New York Times, May 13, 2024, https://www.nytimes.com/2024/05/13/us/solar-storm-tractor-break-nebraska.html, downloaded: July 20th 2024.

[2] European Commission. Smart Cities - Smart Living. Shaping Europe's digital future, 2019, https://ec.europa.eu/digital-single-market/en/smart-cities, downloaded: July 7th 2019.

[3] Department for Business innovation & Skills. Global Innovators: International Case Studies on Smart Cities Smart Cities Study (Bis research paper no. 135), 2013, https://www.gov.uk/government/publications/smart-cities-international-case-studies-global-innovators, downloaded: July 5th 2019.

[4] Komninos, N. The age of intelligent cities: smart environments and innovation-for-all strategies (First Edition). Routledge, Taylor & Francis Group, New York, 2015.

[5] Picon, A. Smart Cities: A Spatialised Intelligence. Wiley, Chichester, England, 2015.

[6] Middleton, S. E; Middleton, L; Modafferi, S. Real-time crisis mapping of natural disasters using social media. IEEE Intelligent Systems, 29(2):9–17, 2014.

[7] Grasic, V; Kos, A; Mileva-Boshkoska, B. Classification of incoming calls for the capital city of Slovenia smart city 112 public safety system using open Internet of Things data. International Journal of Distributed Sensor Networks, 14(9), 2018.

[8] Grašič, V; Robnik, A. Increasing resilience and resistance of Smart Cities by forecasting the number of incoming emergency calls based on IoT open data. EENA Conference. 27-29 April 2022, Marseille, France, 2022

[9] Miranda, R; Ramos, V; Ribeiro, E; Rodrigues, C; Silva, A; Duraes, D; Analide, C; Abelha, A; Machado, J; Crowdsensing on Smart Cities: A Systematic Review. 2023. doi: 10.1007/978-3-031-22419-5_9, downloaded: July 20th 2024.

[10] Suhag, D; Jha, V. A comprehensive survey on mobile crowdsensing systems. Journal of Systems Architecture, Vol. 142. 102952, 2023, https://doi.org/10.1016/j.sysarc.2023.102952, downloaded: July 20th 2024.

[11] HDGM. Modeling Earth's Geomagnetic Fields. National Oceanic and Atmospheric

Administration (NOAA), May 29, 2019, Updated July 19, 2024, https://www.ncei.noaa.gov/news/HDGM, downloaded July 20th 2024.
[12] Cires. Introduction to geomagnetism. Cooperative Institute for Research in Environmental Sciences (Cires), University of Colorado, 2024, https://geomag.colorado.edu/magnetic-field-overview, downloaded: July 20th 2024.
[13] Olsen, N; Hulot, G; Sabaka, T. Sources of the Geomagnetic Field and the Modern Data That Enable Their Investigation. In: Freeden, W; Nashed, M.Z; Sonar, T. (eds) Handbook of Geomathematics. Springer, Berlin, Heidelberg, doi: 10.1007/978-3-642-01546-5_5, downloaded: July 20th 2024.
[14] Hapgood, M. (2017a). Space Weather. IOP Publishing, Bristol, 2017, doi: 10.1088/978-0-7503-1372-8, downloaded: September 20th 2024.
[15] Nicoll, K.A. (2014), Space weather influences on atmospheric electricity. Weather, 69: 238-241, 2014, doi: 10.1002/wea.2323, downloaded: July 20th 2024.
[16] Daglis, I. A; Thorne, R. M; Baumjohann, W; Orsini, S. The terrestrial ring current: Origin, formation, and decay. Reviews of Geophysics, 37(4), 407-438, 1999, doi: 10.1029/1999RG900009, downloaded: July 20th 2024.
[17] INTERMAGNET. International Real-time Magnetic Observatory Network, https://intermagnet.org/, downloaded: September 20th 2024.
[18] Alken, P; Thébault, E; Beggan, C.D. et al. International Geomagnetic Reference Field: the thirteenth generation. Earth Planets Space 73, 49 (2021), doi: 10.1186/s40623-020-01288-x, downloaded: July 20th 2024.
[19] CrowdMag. NCEI Geomagnetism Group. National Oceanic and Atmospheric Administration (NOAA), https://www.ncei.noaa.gov/products/crowdmag-magnetic-data, downloaded: July 20th 2024.
[20] Duma, G. Results of geomagnetic measurements in Slovenia. Internal raport, Wien, Zentralanstalt für Meteorologie und Geodynamik.1998
[21] Čop, R. Geomagnetne meritve na ozemlju Republike Slovenije, https://fgg-web.fgg.uni-lj.si/sugg/referati/2020/SZGG_IAGA_25_let_v_Sloveniji-Cop.pdf, downloaded: September 20th 2024.
[22] Berger, J; Duma, G; Leichter, B; Geomagnetic Measurements in Slovenia 1998 Performed during the Geomagnetic Survey of Austria 1995-1998. Presented on International scientific conference on Magnetism-Geomagnetism-Biomagnetism MGB – 2008. Sezana: VSS Sezana, 2008.
[23] Grašič, V. Real-time analysis of status and changes of electromagnetic field using CrowdMag application for Smart Environment, DIEC 2024, Book of proceeding 7, 7 (2024), Tuzla, https://ipi-akademija.ba/file/diec-7-online/211, downloaded: September 20th 2024.
[24] Čop, R; Deželjin, D; Mihajlović, J. S; Kosovac, P. Preliminary Measurements of Geomagnetic-field Variations in Slovenia, Elektrotehniški vestnik, 78 (3), 96–101, (2011).
[25] GeoM-v1. GeoM-v1 SafeCity112, https://github.com/SafeCity112/GeoM-v1, downloaded: September 20th 2024.
[26] Phyphox. phyphox Sensor Database, https://phyphox.org/sensordb/, downloaded: July 20th 2024.
[27] WMM. World Magnetic Model (WMM). National Centres for Environmental Information (NCEI). National Oceanic and Atmospheric Administration (NOAA), https://www.ncei.noaa.gov/products/world-magnetic-model, downloaded: July 20th 2024.

Enhancing Transparency and Efficiency in Small Size Loans Using Blockchain: A Shared Research Ledger Solution

Dániel Kovács [1], Bálint Molnár [2], Viktor Weininger [1]
[1] Abris Kft. (LLC.), Montevideo u. 6.
1037 Budapest, Hungary,
{daniel.kovacs, viktor.weininger}@abrisconsult.com
[2] Eötvös Loránd University(ELTE),
Information Systems Department, Pázmány Péter 1/C,
1117 Budapest, Hungary, molnarba@inf.elte.hu

Abstract. *This paper explores the integration of the Shared Research Ledger (SRL), a blockchain-based research ledger, with the small-size loan application process to enhance transparency, efficiency, and trust. By leveraging blockchain technology, this solution aims to address key challenges in the lending sector, such as loan stacking, data privacy, and inefficient loan processing. The proposed system employs advanced encryption, zero-knowledge proofs, and smart contracts to ensure secure, real-time data sharing while protecting borrower privacy. This study discusses the implementation framework, benefits, challenges, and potential impacts on the financial ecosystem. Industrial research and development follows the proper research methodologies and applies adequate business process modeling and pattern-based approaches exploiting Service Oriented Computing.*

Keywords. Blockchain, Business Process Modeling, Lending, Transparency, Crowd financing, Loan management

1 Introduction

The lending industry is encountering significant challenges in ensuring transparency, explainability, and trustworthiness, especially in the small-size loan sector. These challenges are further amplified by the use of AI or machine learning tools.

Loan stacking, where borrowers take multiple loans from different lenders simultaneously, poses a substantial risk to financial institutions, increasing default rates and operational costs. Moreover, traditional lending processes are often slow and bureaucratic, further exacerbating the issue [15].

Blockchain technology offers a promising solution to these problems by providing an immutable, transparent ledger for all loan-related activities. The blockchain can live up to the expectations through exploiting the model checking and semi-formal modeling of smart contracts, and auditing the distributed application [18, 8]. The Shared Research Ledger (SRL) system, initially developed for early publication registration in research

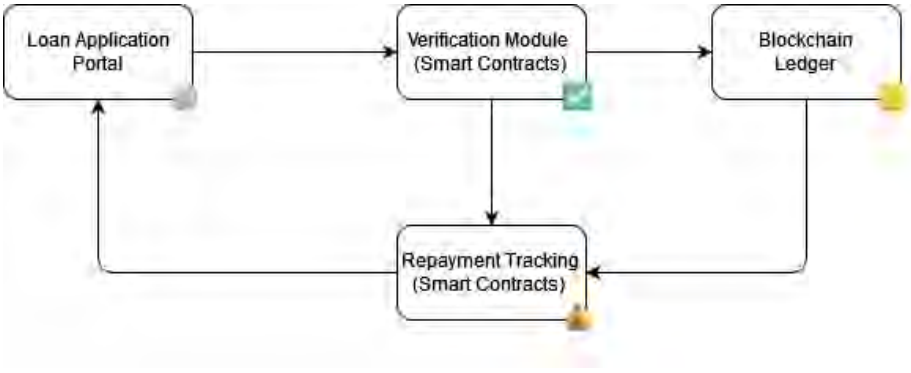


Figure 1: Overview of the Architecture

[35], can be adapted to enhance the small-size loan application process. This paper examines the potential of an SRL-based solution to improve loan processing and provide the opportunity for explainability and interpretability.

2 Background and Motivation

Current Challenges in Small-Size Lending. Small businesses and individuals often face significant barriers in accessing credit due to stringent requirements and lengthy approval processes. The prevalence of loan stacking, where borrowers take advantage of the slow data-sharing processes among lenders, further complicates the lending landscape [20]. This practice not only increases the risk of default but also undermines the trust between lenders and borrowers.

3 Related works

The literature review articles by Casino et al. and Gorkhali et al. summarize important findings on the increasing adoption of blockchain technology in the financial industry, influenced by the emergence of Bitcoin. They highlight the continued progress in creating regulations and structures to improve efficiency and productivity, and provide guidance for further studies in this area.[6, 12]. Viriyasitavat et al.'s study delves into the application of Blockchain technology (BCT) to optimize Business Process Management (BPM) in the context of Industry 4.0, concentrating on digitalization and automating processes while ensuring trust and overseeing digital assets. Their work presents an automated BPM structure utilizing BCT (Blockchain Technology) to check business reliability, assess Quality of Services, and facilitate service selection and integration in open business settings to enhance effectiveness and dependability. [30]. Kherbouche et al.'s research investigates how blockchain technology can improve business process modeling and validation in the financial industry, with a specific focus on insurance claims. This approach has been validated using Petri nets and relational logic, leveraging the Alloy framework. Additionally, to improve workflow coherence, they utilized the services of the YAWL workflow modeling framework. The Alloy model checking tool was employed in conjunction to verify and validate the correctness of business processes. The paper discusses existing methodolo-

gies and suggests future research directions for utilizing blockchain in financial business process modeling [19]. Molnár et al.'s study investigates the potential of blockchain technology to enhance business process optimization in competitive markets by enhancing efficiency, effectiveness, and secure data exchange. With a specific focus on the banking and insurance industries, the paper seeks to offer tailored management models based on blockchain, evaluate the challenges and prospects of implementation, and explore the utilization of the loan process as a web service to support organizations in innovating and adjusting their operations.[22]. Bransia et al. thoroughly examine the functionalities, adoption barriers, and computational processes of two potential blockchain technology environments for financial applications: the Flow and Cardano blockchains. They delve into aspects such as regulatory compliance, security, and scalability, while also exploring data analytics, security enhancement, and readiness for quantum computing through computational methods. The findings highlight the unique strengths and areas of potential application of both platforms in various areas of finance. Understanding how to address these challenges through the examples of Flow and Cardano is critical for stakeholders in the rapidly evolving blockchain landscape. The paper provides insights that decision-makers, academics, and developers can use to gain a deeper understanding of the application of blockchain in the field of microfinance and insurance [5].

4 Methodology

The study in question centered around a software experiment designed to create a new service for a bank in the FinTech space. The main research question examined the effectiveness of blockchain technology in supporting highly regulated businesses in the financial sector. There were concerns about whether the system could deliver on the promises of blockchain, including maintaining trust, traceability, and reliable data management, as well as compliance with existing business regulations. Following the design science research methodology, a design activity was carried out, leading to the meticulous development of the system and software architecture. This design employed a pattern-based strategy that integrated particular blockchain patterns. Additionally, workflow patterns were utilized to outline the business processes in detail [14, 32]. The approach involved leveraging a design pattern-based framework, which included blockchain patterns [33], along with workflow patterns to thoroughly define the business processes [26]. A prototype demonstration was then assessed by potential users, in alignment with the software case study methodology [24].

5 SRL-Based Solution Framework

5.1 Blockchain as a Solution

Blockchain technology, with its decentralized and immutable nature, offers a robust solution to these challenges. By recording all loan transactions on a blockchain, financial institutions can ensure that all activities are transparent and verifiable. Smart contracts can automate various aspects of the loan process, reducing the need for intermediaries and speeding up approval and disbursement times [13].

5.2 Research issues for an Implementation Strategy

Blockchain technology can establish an immutable record of all loan-related activities, including document generation, data collection, and decision-making records. This facilitates automated processing of input and output data. Smart contracts can streamline the loan application, verification, and approval processes, ensuring all requirements are satisfied before the loan is issued. Each blockchain transaction can represent the steps of these procedures.

We can register the start of loan applications on the SRL system, making the entire process transparent from the beginning (refer to Fig. 1 for an overview of the architecture). This early registration database, the *SRL* should be accessible by interested parties, banks, credit agencies, crowd-funding, crowd-financing providers, and other financial institutions in private cloud computing and permissioned blockchain environments. A permissioned blockchain was developed where only authorized entities, such as financial institutions and regulators, can access the data. This ensures data privacy while maintaining transparency and accountability. This helps prevent borrowers from applying for multiple loans at the same time by providing real-time updates visible to all participating lenders.

When a borrower applies for a loan or financial resources, the application is promptly registered on the SRL system. This initial registration includes important details such as the borrower's identity (encrypted), loan amount, and application timestamp.

As the application for a loan or financial resources goes through verification and approval stages, the *SRL* system updates the blockchain distributed ledger in almost real-time. This ensures that all participating lenders have access to the most recent information about the borrower's loan applications.

The SRL system provides a clear and real-time record of loan applications, which helps to prevent borrowers from applying for multiple loans at the same time without being noticed. Lenders can check if a borrower has already applied for a loan elsewhere, which reduces the risk of taking on too much debt.

The developed solutions exploit the advanced cryptography customized for the *SRL*. For this reason, the system employs advanced encryption techniques to protect sensitive borrower information. Use zero-knowledge proofs to verify data without revealing the underlying information, ensuring compliance with GDPR, CPA, and other data protection regulations [29, 23, 10] .

6 Case Study and Evaluation

6.1 Case Study: Small Business Lending

A pilot implementation of the *SRL-based* solution was conducted with several small businesses applying for loans through participating financial institutions. The system successfully registered loan applications on the blockchain, verified borrower information using zero-knowledge proofs, and automated the approval and disbursement processes using smart contracts [29].

The challenges faced by small businesses in accessing credit, such as lengthy approval processes and lack of transparency, are well-documented [17]. The *SRL-based* solution addresses these challenges by providing a transparent, immutable record of all loan-related activities, significantly reducing processing times and enhancing trust among

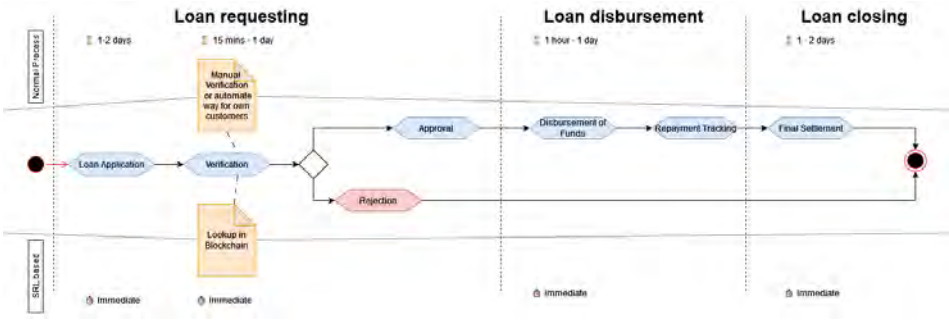


Figure 2: Loan Request Activity Diagram

stakeholders (see Fig. 2). Furthermore, in accordance with the obligation to provide basic information, the system can support the generation of explanations of the reasons for refusals.

6.2 Discussion

The industrial research and development had two facets: business/finance and IT. For this reason, we planned a Business Case Study and a Software Case Study [24]. We investigated the feasibility of the proposed system by the method of research through design. Besides the issues raised by the industrial partner, the business research in the international state-of-the-art small business lending was taken into account [17, 16] The pilot implementation demonstrated several key benefits:

- **Reduced Processing Times**: Loan processing times were significantly reduced, with most applications approved and disbursed within 24 hours.
- **Enhanced Transparency**: All loan transactions were recorded on the blockchain, providing a clear and verifiable audit trail.
- **Improved Trust**: Borrowers and lenders reported increased trust in the system due to its transparency and security features.

7 Recommended Blockchain Architecture for SRL

To implement the Shared Research Ledger (SRL) solution effectively, a blockchain architecture that balances transparency, security, and efficiency is crucial. The key blockchain architecture that we have considered in technical implementation was Hyperledger Fabric [3]. We investigated alternative blockchain technologies, namely Ethereum, Cardano/Marlowe, besides Hyperledger Fabric [21, 7].

Hyperledger Fabric can provide the services as follows: (a) Allows customization of components like membership services and consensus algorithms in a modular architecture, (b) makes the automation of loan applications possible through Smart Contracts, and then the verification, approval, and repayment processes, and (c) offers secure channels for confidential exchanges among designated parties, guaranteeing that sensitive information is shared exclusively with authorized individuals

We carried out a software experiment to build up the customized version of SRL on the basis of Hyperledger Fabric technology. Our research and development team (i) set up the network with the required nodes and established membership services to oversee the participants, (ii) created and implemented smart contracts *(chain code)* for automated management of loan application workflow, from registration to approval, as well as verifying user details and tracking loan repayment history, (iii) integrated with existing banking systems for seamless data flow and real-time updates on the blockchain in a test environment, and (iv) conducted thorough tests to guarantee the solution's security, performance, and compliance prior to its production deployment

7.1 Design Patterns in Blockchain Solutions for SRL

Several design patterns in blockchain solutions can be effectively applied to the SRL (Shared Research Ledger) solution. These patterns help address common challenges and optimize the functionality and performance of blockchain systems [34].

Small-size lending is a complex business process internally in a financial institution so the complicated interactions can be mapped rigorously by exploiting the workflow and business process patterns [1, 2]. The blocks in a blockchain can contain only a limited size of information therefore there is a need to store data off-chain. The data interchange between the blockchain and other processes and storage areas can be realized through Service Oriented Computing, Web Service, and APIs (Application Programming Interfaces) [9]. The activities of business processes contain the links to Web services or APIs that are modeled by Business Process Modeling Notation[31].

7.2 The proposed experimental architecture

Our research and development team used the design patterns in a disciplined way. (1) critical loan application data is stored *On-Chain*, i.e. transaction history, and verification statuses on the blockchain to ensure transparency and immutability., (2) large files, detailed documents, and sensitive personal information in a secure *Off-Chain* storage solution are kept and linked to the blockchain via cryptographic hashes, (3) *Multi-Signature (Multisig) Transactions* were employed for loan approval processes, where a loan must be approved by multiple parties (e.g., lender, borrower, and possibly a regulatory body) before funds are disbursed, (4) *Smart contracts* are used to automate the verification, approval, and repayment processes for loans. Conditions such as credit checks and repayment schedules can be coded into the smart contracts, (5) *Oracle pattern* is a service that provides external data to smart contracts, enabling them to interact with real-world data outside the blockchain. Oracle patterns were integrated to fetch external data such as credit scores, KYC (Know Your Customer) information, and regulatory compliance checks. This ensures that the blockchain has access to up-to-date and accurate information for decision–making, and (6) *Interoperability pattern* allows different blockchain networks to communicate and share data with each other. This pattern is essential for integrating with existing systems and other blockchain networks. The interoperability pattern ensures that the SRL solution can interoperate with other financial systems and blockchains. This can be achieved using cross-chain communication protocols, allowing seamless integration and data sharing.

8 Summary

The integration of blockchain technology through the Shared Research Ledger (SRL) system offers a transformative approach to the small-size loan industry and can extend its benefits to the crowdfunding domain. By leveraging the transparency, immutability, and efficiency of blockchain, the SRL-based solution addresses key challenges such as loan stacking, data privacy, and slow verification processes that plague traditional lending and crowdfunding models [4].

8.1 Compliance by design

The implementation strategy involving automated processes through smart contracts, real-time updates on the SRL, and advanced encryption techniques ensures that both lenders and borrowers benefit from a secure and efficient system. The pilot case study demonstrated significant improvements in processing times, transparency, and trust among stakeholders, indicating the potential for widespread adoption.

In the context of crowdfunding, the SRL-based solution can enhance transparency and trust by providing an immutable record of all transactions and project updates. This transparency can prevent fraudulent activities and reassure backers about the legitimacy and progress of projects. Additionally, smart contracts can automate the distribution of funds based on predefined milestones, ensuring that funds are used as intended and reducing operational costs [11, 27].

Furthermore, by adhering to data protection regulations such as GDPR and CPA, the SRL system not only enhances operational efficiency but also safeguards participant privacy, a crucial aspect in today's regulatory environment. The use of zero-knowledge proofs and permissioned blockchain ensures compliance without compromising the integrity of the data.

8.2 Scientific contribution

The goal of the project was to conduct a software experiment through industrial research and development. The primary research question focused on whether current technologies could facilitate the creation of an IT system to support complex business processes. The development and implementation phases helped to clarify the feasibility of the proposed solution. Using business process management and modeling, we established a business and IT architecture. We applied Service Oriented Computing to align the requirements of the business processes with the services offered by the underlying blockchain technology [28]. Our approach was based on business process modeling, using the Business Process Modeling Notation for process representation [31]. This choice was made because the representation allows for model checking [18]. We integrated workflow and blockchain patterns into a service-oriented, API-based framework [1, 33, 9]. A tailored approach to the distributed ledger was developed in the project, which was named the Shared Research Ledger (SRL). The development result was presented to the representatives of the funding agency and the industrial partner. The prototype system was assessed according to the Software Case Study Methodology [24]. A LIKERT questionnaire was created to survey the impression of end-users and experts of IT architectures [25].

8.3 Limitations of the Research

During our research and development, we found that there was a lack of comprehensive auditing and validation methods for decentralized applications. This is a limitation of our study and indicates areas for future work in cybersecurity, reliability, and trustworthiness of the solution (EEA2023a).

Our research goal was to create a "proof-of-concept" for the proposed system. While experts in the banking industry and emerging crowd-funding companies have accepted the ideas and technological solutions, full implementation will require significant effort and business process improvement in the system's administration. Additionally, the system should incorporate an optimization and match-making process to connect requestor and investors, minimizing operational risk and yielding reasonable results by applying recent Data Science technologies (TanZhengZhuEtAl2017a).

9 Conclusion

The industrial research and development project demonstrated the feasibility and viability of the proposed Enterprise and IT architecture.

In conclusion, the SRL-based solution represents a significant advancement in both the lending and crowdfunding industries, offering a robust framework for managing small-size loans and crowdfunding campaigns. *Future research and development* should focus on scaling this solution, auditing, validating, and integrating it with existing financial systems. Furthermore, the opportunities should be explored for their applicability to other areas of finance, e.g in insurance.

Data Availability Statement:

The data and code presented in this study are available on request from the authors [35].

10 Acknowledgments

The paper was supported by the "Application Domain-Specific Highly Reliable IT Solutions" project, implemented with the support provided by the National Research, Development and Innovation Fund of Hungary, financed under the Thematic Excellence Programme TKP2021-NVA-29 (National Challenges Subprogramme) funding scheme, furthermore, NKFIH-KFI-2018.1.1.2-KFI-2018-0056, and by the COST Action CA19130.

References

[1] Wil Aalst. *Workflow management : models, methods, and systems*. MIT Press, Cambridge, Mass, 2002.

[2] American Productivity Quality Center. APQC Process Classification Framework (PCF) - Cross Industry - Excel Version 7.3.1, August 2024. [Online; accessed 18. Aug. 2024].

[3] Elli Androulaki, Artem Barger, Vita Bortnikov, Christian Cachin, Konstantinos Christidis, Angelo De Caro, David Enyeart, Christopher Ferris, Gennady Laventman, Yacov Manevich, Srinivasan Muralidharan, Chet Murthy, Binh Nguyen, Manish Sethi, Gari Singh, Keith Smith, Alessandro Sorniotti, Chrysoula Stathakopoulou, Marko Vukolić, Sharon Weed Cocco, and Jason Yellick. Hyperledger fabric: a distributed operating system for permissioned blockchains. In *Proceedings of the Thirteenth EuroSys Conference*, EuroSys '18. ACM, April 2018.

[4] CMR Berkeley. Unveiling and brightening the dark side of crowdfunding. *California Management Review*, 2021.

[5] Ayda Bransia, Bálint Molnár, and Simon Thompson. Comparative analysis of flow and cardano blockchains: Navigating adoption challenges, computing techniques and implications for the blockchain landscape. In *Intelligent Computing*, pages 665–670. Springer Nature Switzerland, 2024.

[6] Fran Casino, Thomas K Dasaklis, and Constantinos Patsakis. A systematic literature review of blockchain-based applications: Current status, classification and open issues. *Telematics and informatics*, 36:55–81, 2019.

[7] Chris Dannen. *Introducing Ethereum and Solidity: Foundations of Cryptocurrency and Blockchain Programming for Beginners*. Apress, 2017.

[8] Enterprise and Ethereum and Alliance and Inc. EEA EthTrust Security Levels Specification version 2, December 2023. [Online; accessed 17. Aug. 2024].

[9] Thomas Erl. *Service-oriented architecture : analysis and design for services and microservices*. Prentice Hall, Service Tech Press, Boston, 2017.

[10] European Union. General Data Protection Regulation (GDPR). Technical report, European Union, 2018.

[11] FasterCapital. Online legal and ethical issues: Legal Challenges in Crowdfunding Campaigns - FasterCapital, April 2024. [Online; accessed 20. Jun. 2024].

[12] Anjee Gorkhali, Ling Li, and Asim Shrestha. Blockchain: A literature review. *Journal of Management Analytics*, 7(3):321–343, 2020.

[13] Gordon Gottsegen. Spring Labs raises $23M to stop loan fraud with blockchain technology | Built In Chicago, June 2019. [Online; accessed 14. Jun. 2024].

[14] Alan Hevner and Samir Chatterjee. *Design Research in Information Systems*. Springer US, 2010.

[15] Christian Hugo Hoffmann. Blockchain use cases revisited: Micro-lending solutions for retail banking and financial inclusion. *Journal of Systems Science and Information*, 9(1):1–15, February 2021.

[16] Stefan Hunziker and Michael Blankenagel. *Research Design in Business and Management: A Practical Guide for Students and Researchers*. Springer Fachmedien Wiesbaden, 2024.

[17] Julapa Jagtiani, Catharine Lemieux, et al. Small business lending after the financial crisis: A new competitive landscape for community banks. *Economic perspectives*, 40(3):1–30, 2016.

[18] Meriem Kherbouche and Bálint Molnár. Formal model checking and transformations of models represented in uml with alloy. In *Modelling to Program: Second International Workshop, M2P 2020, Lappeenranta, Finland, March 10–12, 2020, Revised Selected Papers 1*, pages 127–136. Springer, 2021.

[19] Meriem Kherbouche, Galena Pisoni, and Bálint Molnár. Model to program and blockchain approaches for business processes and workflows in finance. *Applied System Innovation*, 5(1):10, January 2022.

[20] Deepak Kumar, B.V. Phani, Naveen Chilamkurti, Suman Saurabh, and Vanessa Ratten. Filling the sme credit gap: a systematic review of blockchain-based sme finance literature. *Journal of Trade Science*, 11(2/3):45–72, September 2023.

[21] Pablo Lamela Seijas and Simon Thompson. *Marlowe: Financial contracts on blockchain*, pages 356–375. Springer International Publishing, 2018.

[22] Bálint Molnár, Galena Pisoni, Meriem Kherbouche, and Yossra Zghal. Blockchain-based business process management (bpm) for finance: the case of credit and claim requests. *MDPI Smart Cities*, 6:1254–1278, 2023.

[23] State of California Department of Justice. California consumer privacy act (ccpa). Technical report, State of California Department of Justice, 2020.

[24] Per Runeson, editor. *Case study research in software engineering*. Wiley, Hoboken, NJ, 2012. Includes bibliographical references (p. 221-223) and index. - Description based on print version record.

[25] Thomas Rusch, Paul Benjamin Lowry, Patrick Mair, and Horst Treiblmaier. Breaking free from the limitations of classical test theory: Developing and measuring information systems scales using item response theory. *Information & Management*, 54(2):189–203, March 2017.

[26] Nick Russell, Wil M. P. van der Aalst, and Arthur H.M. ter Hofstede. *Workflow Patterns*. MIT Press Ltd, 2016.

[27] Md Nazmus Saadat, Syed Abdul Halim, Husna Osman, Rasheed Mohammad Nassr, and Megat F. Zuhairi. Blockchain based crowdfunding systems. *Indonesian Journal of Electrical Engineering and Computer Science*, 15(1):409, July 2019.

[28] Munindar P Singh and Michael N Huhns. *Service-oriented computing: semantics, processes, agents*. John Wiley & Sons, Chichester, England, 2005. Use copy Restrictions unspecified star MiAaHDL.

[29] Balázs Ádám Toldi and Imre Kocsis. Blockchain-based, confidentiality-preserving orchestration of collaborative workflows. *Infocommunications journal*, 15(3):72–81, 2023.

[30] Wattana Viriyasitavat, Li Da Xu, Zhuming Bi, and Arnon Sapsomboon. Blockchain-based business process management (bpm) framework for service composition in industry 4.0. *Journal of Intelligent Manufacturing*, 31(1):1737–1748, 2020.

[31] Stephen A. White and Derek Miers. *BPMN modeling and reference guide*. Future Strategies Inc., Lighthouse Point, Fla., 2008.

[32] Roel J. Wieringa. *Design Science Methodology for Information Systems and Software Engineering*. Springer Berlin Heidelberg, 2014.

[33] Wei Xu. Toward human-centered ai: a perspective from human-computer interaction. *interactions*, 26(4):42–46, 2019.

[34] Xiwei Xu, Ingo Weber, and Mark Staples. *Architecture for Blockchain Applications*. Springer International Publishing, 2019.

[35] Ábris Informatikai és Tanácsadó Korlátolt Felelősségű Társaság. Projekt nkfih-kfi-2018.1.1.2-kfi-2018-0056 a blokklánc technológiai megoldások vizsgálata és fejlesztése nyugdíjbiztosítási információk és a tudományos publikációk tárolásának területén (Testing and developing blockchain technology solutions for the storage of pension information and scientific publications), September 2019. Accessed:2021-02-12.

Harnessing Big Data: Advantages and Challenges in Macroeconomic analysis and Economic Forecasting

Milica Stanković, Gordana Mrdak
Academy of Applied Technical and Preschool Studies
Filipa Filipovića 20, 17000 Vranje, Serbia
{milica.stankovic, gordana.mrdak}@akademijanis.edu.rs

Abstract: *Big data transformed macroeconomic analysis and economic forecasting by providing real-time information and better economic insight. The aim of the paper is to emphasize the benefits of big data in increasing the accuracy of forecasts, making better financial decisions and monitoring of economic indicators. Also, the paper outlines challenges that arise in using big data, including data privacy and security, data quality, technical infrastructure and professional skills needed to analyze big data. Overcoming these challenges will ensure big data to become increasingly important tool for policymakers and economists to achieve sustainable economic growth and stability. The study emphasizes the need for cooperation between data scientists, policymakers, and other stakeholders to maximize the potential of big data and minimize risks.*

Key Words*: Big Data***,** *Macroeconomic Analysis***,** *Economic Forecasting***,** *Policy Making*

1 Introduction

Big data refers to large amounts of data, both structured and unstructured. Integrating big data with official statistics and economic analysis has numerous opportunities and challenges for statisticians and users of statistical information. These large datasets can be very useful in generating new indicators and enabling faster predictions based on official statistical data. With the rapid growth of information, further expedited by the Internet, we entered the era of big data. Big data has great influence both on microeconomic and macroeconomic analyses. At the micro level, big data is very useful for industry research and business applications. At the macro level, international institutions, organizations and governments use big data in economic forecasts, policy analysis and future policy decision-making. The aim of this paper is to emphasize potential advantages and challenges of big data in economic forecasting and macroeconomic analysis.

2 Big Data: Benefits and Challenges

Big data represents a very important step in data collection and analysis, where innovative real-time analysis can give major benefits for economic and financial analysis. Although big data has a lot of benefits, the opportunities for using it differ significantly across countries, bearing in mind that different countries are on different level of technological development. Unlike traditional statistical information, which is collected with specific

purposes, big data is generated as a by-product of business and administrative systems, social media, and the internet of things [1]. Often, the 3Vs are related to big data: high volume, high velocity, and high variety. More recently, other characteristics, such as veracity and volatility, have been added.

Big data got attention from both private and the public sectors. In the private sector, big data is used to gain a competitive advantage by collecting data for marketing insights and customized interactions with customers as well as dynamic content management. In the public sector, big data is applied to improve operational efficiency. Central banks are investigating the application of big data in predicting macroeconomic indicators and facilitating financial analyses [1]. The main benefit of big data is the potential to deliver faster and more precise economic insights, especially in areas such as financial markets and pricing information. One of the significant advantages of big data is that it can track core economic indicators in almost real-time, enabling forecasting, which makes economic monitoring more effective. By providing real-time information, big data can help policymakers assess the economic situation of individual economies and the global economy, so it can contribute to avoiding financial crises [2].

Despite its benefits, big data presents numerous challenges. Among the primary concerns are the assurance of data integrity and the consistent availability of data sources. Furthermore, a critical challenge lies in the governance of privacy, confidentiality, and cybersecurity threats. Given that big data often consists of sensitive information, it is very important to protect this data from unauthorized access and cyberattacks [3] [4]. Furthermore, other costs related to big data, such as those associated with purchasing and processing, can also be quite significant, including software licenses and all equipment used in the process. Converting big datasets into valuable inputs for official statistics requires multidisciplinary teams with an appropriate level of technical capabilities (i.e., information technology, statistics instrumentation, analysis, and domain experience), so it is crucial to emphasize the important of cooperation between statistical agencies and diverse professionals from this field [1].

3 Big Data in Economic Forecasting and Macroeconomic Analysis

Big data applications in macroeconomic forecasting provide a lot of benefits but, at the same time, significant challenges. Big data is particularly useful in tackling the issue of time delays experienced in administrative statistical data, hence enabling quicker predictions and analysis in a wide range of areas such as national accounts, external sectors and finance statistics [1]. The development of big data technologies has significantly impacted economic forecasting and macroeconomic analysis [5]. Traditionally, economic forecasting mainly relied on econometric models that used historical data and economic theories to predict important economic variables such as GDP growth, inflation, and employment [6]. Big data has changed the face of economic forecasting by providing better and accurate details, apart from enabling the real-time monitoring of economic conditions [7] [8]. Through the utilization of diverse data sources, big data facilitates a more holistic understanding of economic activities, thereby acquiring information that was formerly unattainable [9].

With the use of big data, decision-making has become quite dynamic, so we can have immediate access to relevant data for informed and expedient decision-making processes [10]. This greatly impacts how economic activity is organized today, since large discrete data sets and asymmetrical information present problems for traditional market structures. The shift towards data-driven policymaking serves to better target and make efficient

resource allocation, minimizing waste while attaining maximum impact [11]. Continuous monitoring and valuation of the policy output allow adjustments at the right time to relevant and effective policies. Big data offers significant benefits for economic forecasting and macroeconomic analysis by enhancing accuracy, timeliness, comprehensiveness, and predictive capabilities. However, challenges related to data privacy, quality, ethics, and technical infrastructure must be addressed to fully realize its potential [12]. By addressing these challenges, big data can enhance economic stability and growth, making it an indispensable tool for future economic analysis and policy making [13]. The benefit of big data is set to significantly enhance economic forecasting and the formulation of policies, facilitating ongoing advancements in economic strategies and decision-making methodologies. However, to get its maximum potential, considerable investments in data governance, technological infrastructure, and skill enhancement are essential [14].

It is important to point out the existence of data management frameworks that have been developed to protect personal data and resolve privacy problems. The GDPR (General data protection regulation) in the European Union sets the gold standard because strict guidelines are followed in data collection and processing ensures. Besides such frameworks, blockchain and other decentralized data architectures have appeared as solutions which may respond to the improvement of data privacy [12]. The complexity of big data points out the importance of integrating deep learning methods into enhancing data analysis. Techniques such as RNN (recurrent neural networks) and CNN (convolutional neural networks) provide advanced capabilities to process huge sets of data and improve accuracy in macroeconomic forecasting [6]. Big data analytics heavily relies on the success analytics professionals, so one of the biggest challenges is lack of professionals who have knowledge and skills in this field. Capacity building programs can help to bridge the gap, particularly in regions with limited technological progress. It is very important to mention the collaborative initiatives between academic institutions and the private sector in skills development to enhance the adoption of big data technologies in macroeconomic analysis.

Significant gaps between countries' relative abilities to exploit big data, especially between advanced economies and developing countries, remain an issue. Developing countries lack technical infrastructures and human skills. In the way of bridging that gap, technology-sharing agreements and investment in the development of necessary infrastructure have been suggested by several international organizations. Such efforts combined with local capacity building programs will ensure the improvement of big data usage in global macroeconomic policy. For Example, G20's Global Partnership for Financial Inclusion (GPFI) encourages cross-country sharing of technologies related to financial data systems and inclusive finance tools. Similarly, IMF's Capacity Development initiatives as an example of multilateral agreements, have goals to enhance the analytical capacities of developing country central banks using economic forecasting models and big data technologies. With such agreements for cooperation, the advanced economies can share big data infrastructure and best practices in data management and forecasting with the less technologically developed nations to help them enhance macroeconomic planning [15].

A good example of practical use of big data in macroeconomic forecasting is the Now-casting model, which is in use at the European Central Bank. Now-casting refers to the forecasting the present, very near future and very recent past of key economic indicators. The European Central Bank has already been able to exploit big data sources-such as real-time Internet searches and financial market data-and integrate them into its current

forecasting models. ECB has integrated data in real time from various sources and succeeded in coming out with more accurate and timely forecasts of GDP, inflation, and unemployment. Big data proved very useful in forecasting during the 2008 financial crisis when traditional econometric models were late in detecting this sudden fall in economic activities, thus helping policymakers make decisions in changing economic conditions. Big data can, hence, support macroeconomic stability through timely insight provision to policymakers for better decisions [2].

Table 1.: Using Big Data for Economic Forecasting and Macroeconomic Analysis: Main Benefits, Challenges and Proposed Solutions

Benefits	Challenges	Proposed Solutions
Increased accuracy **Big data provides detailed information, improving the accuracy of economic models and forecasts.**	**Data privacy and security** The collection and analysis of vast amounts of personal data raise significant privacy issues. Protecting data from unauthorized access requires robust security measures and continuous monitoring.	GDPR (General data protection regulation) Blockchain for decentralized security
Predictive capabilities **Integration with machine learning and AI enhances predictive capabilities by identifying complex patterns and relationships in large datasets.**	**Technical and skill barriers** Handling large volumes of data requires scalable infrastructure and advanced technology. There is a high demand for skilled data scientists and analysts.	Investments in technology-sharing Cloud computing Capacity-building programs Collaboration with academia
Timeliness and responsiveness **Big data allows real-time analysis, enabling timely assessment of economic conditions and rapid policy responses.**	**Data quality and reliability** Errors in data collection, inconsistencies from multiple sources, and outdated information can compromise the reliability of analyses. Ensuring high-quality data and timely updates is crucial.	Advanced machine learning RNNs CNNs

4 Conclusion

Big data in use for economic forecasting and macroeconomic analysis represents a big change that has important implications for improving the accuracy, timeliness, and efficacies of economic analyses and forecast. Policymakers and economists can achieve better accuracy in their forecasts from high-resolution data, respond dynamically to newly emerging trends in the economy, and gain sound insight into sectoral and regional conditions. The transformative power of big data is associated with significant challenges that must be addressed to realize all its benefits. Data privacy and security must be ensured to prevent personal information from being misused and to prevent unauthorized access to data. A high degree of data quality and reliability must be maintained to avoid mistakes that may undermine economic forecasts and policy measures. The technical complexities and the skill gaps in harnessing the full power of big data are another challenge. Such challenges can only be overcome through combined effort by data scientists, policymakers, lawyers, and other stakeholders in developing robust measures of data privacy, ensuring accuracy and timeliness of data, handling ethical concerns, making the necessary investments in technical infrastructure, and building skills. Such proactive attention to problems should allow policymakers to unlock the full potential of big data for sustainable economic growth, policy effectiveness, and overall economic stability.

Future research can be directed towards the analysis of data integration methods, especially in developing countries with underdeveloped infrastructure. In addition, the focus should be on ethical issues and privacy concerns, to ensure that big data can be used responsibly in making important macroeconomic decisions. It is difficult to envision economic forecasting without machine learning and artificial, so it will be important to continue understanding how these technologies evolve to further enhance the accuracy of data-driven modeling. The continuous investment in new technologies and data management is a determining factor, as policymakers and economists harness the full power of big data in the interests of fostering sustainable economic growth and stability.

9 References

[1] Hammer, C., Kostroch, D., Quirós, G., STA Internal Group (2017). Big Data: Potential, Challenges, and Statistical Implications, IMF STAFF DISCUSSION NOTE

[2] Banbura, M., Giannone, D., Modugno, M., Reichlin, L. (2013). Now-Casting and the Real-Time Data Flow. Working paper series 1564, European Central Bank, Frankfurt.

[3] Moreno, J., Serrano, M., Fernández-Medina, E. (2016). Main Issue in Big Data Security. Futur. Internet, vol. 8, no. 3, p. 44.

[4] Smith, M., Szongott, C., Henne, B., Von Voight, G. (2012). Big Data Privacy Issues in Public Social Media. In Digital Ecosystems Technologies (DEST), 6th IEEE International Conference, pp. 1–6.

[5] Ikegwu, A.C., Nweke, H.F., Anikwe, C.V., Alo, U.R., Okonkwo, O.R. (2022). Big data analytics for data-driven industry: a review of data sources, tools, challenges, solutions, and research directions. Cluster Computing, 25(5), pp.3343-3387.

[6] Nabipour, M., Nayyeri, P., Jabani, H., Mosavi, A., Salwana, E. (2020). Deep learning for stock market prediction. Entropy, 22(8), p.840.

[7] Kandt, J., Batty, M., (2021). Smart cities, big data and urban policy: Towards urban

analytics for the long run. Cities, 109.
[8] Barlas, A.B., Güler, S., Ortiz, A., Rodrigo, T., (2020). Investment in real time and High definition: A Big data Approach (No. 2013). BBVA Research Working Papers.
[9] Ahmad, T., Zhang, D. (2021). Using the internet of things in smart energy systems and networks. Sustainable Cities and Society, 68.
[10] Shen, T. (2021). Research on economic development strategy based on big data, E3S Web of Conferences.
[11] Bello A. (2022). The Impact of Big Data on Economic Forecasting and Policy Making, International Journal of Development and Economic Sustainability, Vol.10, No.6, pp. 66-89.
[12] Ogbuke, N.J., Yusuf, Y.Y., Dharma, K., Mercangoz, B.A. (2022). Big data supply chain analytics: ethical, privacy and security challenges posed to business, industries and society. Production Planning & Control, 33(2-3), pp.123-137.
[13] Poza, C., Monge, M. (2020). A real time leading economic indicator based on text mining for the Spanish economy. Fractional cointegration VAR and Continuous Wavelet Transform analysis. International Economics, 163, pp.163-175.
[14] Arent, D.J., Green, P., Abdullah, Z., Barnes, T., Bauer, S., Bernstein, A., Berry, D., Berry, J., Burrell, T., Carpenter, B., Cochran, J. (2022). Challenges and opportunities in decarbonizing the US energy system. Renewable and Sustainable Energy Reviews, 169.
[12] GPFI (2023). G20 2023 Financial Inclusion Action Plan, https://www.gpfi.org/publications/g20-2023-financial-inclusion-action-plan

Revolutionizing Pension Systems: The Integration of Blockchain with Business Process Management

Dániel Kovács [1], Bálint Molnár [2], Viktor Weininger [1]
[1] Abris Kft. (LLC.), Montevideo u. 6.
1037 Budapest, Hungary,
{daniel.kovacs, viktor.weininger}@abrisconsult.com
[2] Eötvös Loránd University(ELTE),
Information Systems Department, Pázmány Péter 1/C,
1117 Budapest, Hungary, molnarba@inf.elte.hu

Abstract. *This paper explores the integration of Business Process Management (BPM) with blockchain technology in pension systems, focusing on how these combined technologies can enhance transparency, efficiency, and regulatory compliance. By automating pension-related processes through BPM and leveraging blockchain's decentralized, secure ledger, pension systems can significantly reduce administrative overhead, improve data integrity, and offer stakeholders greater transparency and trust. The paper discusses technical challenges such as interoperability between legacy systems and blockchain platforms, as well as the complexities of ensuring regulatory compliance with data protection laws like GDPR. Additionally, it highlights the importance of addressing operational challenges, such as managing stakeholder engagement and change management during implementation. Through a case study of the Hungarian pension system, this paper demonstrates how BPM and blockchain integration can transform pension operations, streamline workflows, and meet the growing regulatory demands. While the implementation of these technologies presents certain challenges, the potential benefits make the integration of BPM with blockchain a valuable endeavor for pension systems aiming to improve both operational efficiency and stakeholder trust.*

Keywords. Blockchain, Business Process Modeling, Public Administration, Pension Benefit

1 Introduction

Pension systems globally play a critical role in ensuring financial security for individuals during their retirement years. However, these systems are often burdened with inefficiencies, delays, and challenges related to transparency, data integrity, and compliance with regulatory frameworks like the GDPR [29, 30]. In Hungary, as in many other countries, the complexity of managing pension data across decentralized systems, while ensuring the accuracy of contributions and benefit calculations, remains a significant challenge [4].

Blockchain technology, with its decentralized and immutable ledger, offers a solution to many of these issues. By integrating Blockchain with Business Process Management (BPM), we propose a novel approach to optimizing pension management workflows, ensuring both efficiency and regulatory compliance [24]. The combination of these technologies facilitates automation in key processes such as data management, benefit disbursement, and real-time auditing, significantly reducing administrative overhead and improving stakeholder trust.

This paper presents a case study focusing on the Hungarian pension system, highlighting the integration of Blockchain with BPM to address common inefficiencies. We explore how this synergy can streamline pension management through the use of smart contracts, secure data storage, and automated processes. Key challenges such as scalability, regulatory compliance, and system integration are also examined, offering practical insights into the implementation of this solution in a real-world setting [24]. The investigation was realized by research through design. A prototype system was created to demonstrate the feasibility and viability of the proposed solution in a case study, namely in the case of the Hungarian Pension System [19, 33]

2 Related Works

Huang et al. in Ref [21] discuss the use of blockchain technology in a time banking system for mutual assistance in home care. They outline the requirements for a secure digital signature, the design and implementation of a blockchain-based time chain model, and the experimental analysis of transaction throughput. The paper proposes a solution to the current problems in time banking by integrating blockchain technology to ensure security, transparency, and efficiency in the system.

Sarker and Datta in Ref [27] explore the implementation of a blockchain-based digitized pension system, outlining the project's flow and the potential benefits it could bring to the pension industry. The authors present a proposed architecture for the system and discuss challenges related to validating benefits and outcomes due to the limited literature and real-world applications in the pension sector. They also refer to existing blockchain adoption in other industries to support their arguments. The text emphasizes how a blockchain-based pension network can improve the industry and connect all participants on a single digital platform.

Hu et al. in Ref [20] investigate the design and implementation of a smart contract solution based on blockchain access control for managing access control policies within a system architecture. They outline key elements of the smart contract operation process, such as administrators, users, and decision points, and describe their interactions. The smart contract structure includes the Policy Enforcement Point (PEP), Policy Decision Point (PDP), and Policy Administration Point (PAP). The performance testing of the smart contract solution shows its efficiency in processing multiple requests concurrently. This work highlights the importance of smart contracts in implementing access control policies and ensuring secure and reliable access capabilities.

Gong and Zang in Ref [17] explore the application of blockchain technology in for-profit pension institutions. They introduce a two-market model to demonstrate the significance of blockchain in upgrading the service modes of these institutions. The text explores existing service mode upgrade schemes and their defects, proposing new approaches to address the challenges faced by for-profit pension institutions. The paper emphasizes the role of blockchain in improving credibility, service quality, and customer

experience within for-profit pension institutions, stressing the need for service upgrades to meet changing market dynamics and consumer preferences.

Cheng et al. in Ref [3] describe the development of a system called VOLTimebank, which utilizes blockchain technology to manage volunteer service records and facilitate mutual pension services. The system aims to address the challenges faced by traditional time banks, such as the lack of permanent data storage and data-sharing issues. VOLTimebank allows volunteers to earn VOLMedals for their service, which can be used to pay for services from other volunteers. The system also incorporates mechanisms for user participation and governance, such as an electoral process for auditors, and a flowchart to illustrate its operation.

Overall, the reviewed literature shows the wide range of blockchain applications in both the pension sector and related industries. The integration of blockchain into pension systems highlights the potential benefits of transparency, security, and process optimization, though challenges such as validation and real-world implementation remain key topics for further exploration.

3 The Role of BPM in Blockchain Technology

Business Process Management (BPM) plays a vital role in enhancing the operational capabilities of blockchain within pension systems, leading to improved efficiency and effectiveness. BPM tools allow for the modeling, execution, monitoring, and optimization of business processes, ensuring seamless operations within a blockchain-based framework. When integrated with BPM, smart contracts can automate business rules, minimizing errors and boosting operational efficiency [24]. Additionally, BPM ensures regulatory compliance by generating audit trails and compliance reports, which are critical in the management of pension systems [29].

While blockchain technology offers significant advantages, such as transparency, security, and decentralization, the addition of BPM introduces essential layers of process optimization and management, which are particularly valuable for the intricate workflows found in pension systems.

The interaction between BPM and blockchain technology is depicted in the following diagram Fig. 1.

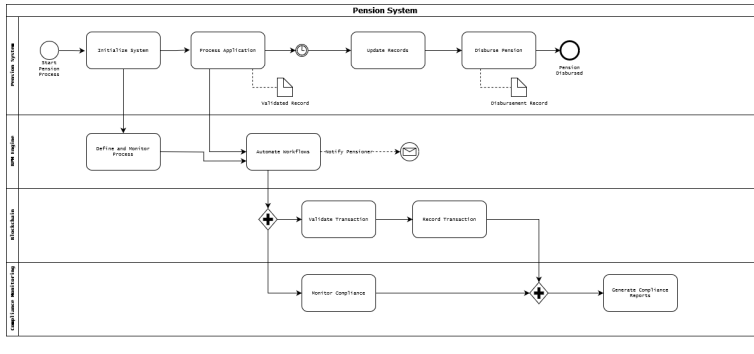


Figure 1: Integration of BPM with Blockchain Technology

This diagram provides an overview of the process, beginning with the input of the retiree's personal and employment information, followed by the retrieval of contribution

records from the blockchain and salary history from a secure database. Pension plan rules are then applied, pension benefits are calculated, the BPM engine verifies the calculations, the final result is recorded on the blockchain, and finally, the retiree is notified of the calculated benefits.

4 Pension Calculation Workflow

Pension systems rely on complex calculations to determine the appropriate benefits for retirees, taking into account factors such as years of service, salary history, contribution amounts, and specific pension plan rules. Without efficient management, these intricate calculations can lead to errors, inefficiencies, and increased administrative costs.

The integration of BPM with blockchain technology can greatly enhance the efficiency and accuracy of pension calculations. BPM tools automate the entire process, ensuring that all applicable rules and conditions are followed consistently throughout the calculation workflow (see Fig. 2).

By automating these processes, the risk of human error is minimized, and calculations can be completed both quickly and accurately. Blockchain technology further strengthens the system by offering a secure and immutable ledger where all transaction records are stored, creating a tamper-proof and transparent history of all pension-related activities. This ensures that all stakeholders can verify the accuracy and trustworthiness of the system [4].

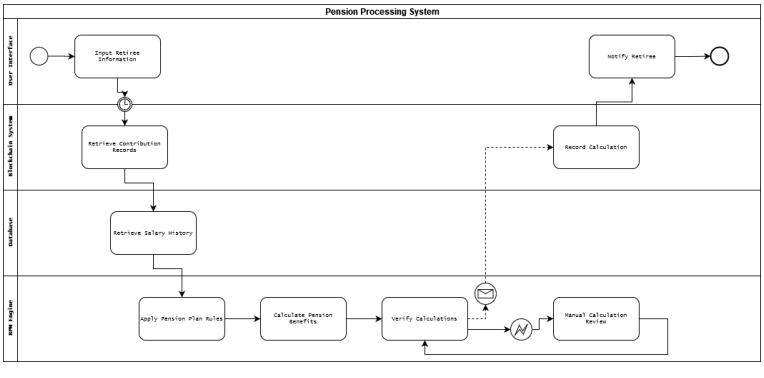


Figure 2: Pension Calculation Workflow

Moreover, the integration of BPM and blockchain not only improves the efficiency of pension calculations but also increases stakeholder confidence. Pensioners, administrators, and regulators alike can access the blockchain to confirm that calculations have been carried out accurately and that all transactions are legitimate. This enhanced transparency helps to reduce disputes and fosters greater trust in the pension system.

5 Data Access Management with PII

In pension systems, the management of data access and the protection of Personally Identifiable Information (PII) are critical concerns. PII includes any data that can identify an

individual, such as names, addresses, Social Security numbers, and financial details. Handling this sensitive information is subject to strict regulatory standards, including the General Data Protection Regulation (GDPR) in the European Union, the California Consumer Privacy Act (CCPA) in the United States, and numerous other global data protection laws [30, 25].

A primary challenge in managing PII within pension systems is ensuring that sensitive data is only accessed by authorized individuals for legitimate purposes. Traditional centralized systems can be susceptible to unauthorized access, data breaches, and misuse. The involvement of multiple intermediaries in processing PII for administrative tasks exacerbates these risks, and compliance with stringent regulatory requirements further complicates data management.

Blockchain technology introduces several features that significantly enhance the security and management of PII in pension systems. By distributing data across a decentralized network, blockchain mitigates the risk of centralized points of failure and makes it more challenging for unauthorized individuals to gain access to the entire dataset. Furthermore, blockchain's immutable ledger ensures that once data is recorded, it cannot be modified or deleted. This immutability guarantees that records are tamper-proof and always available for audit. Additionally, advanced cryptographic techniques can be employed to secure data, ensuring that PII remains protected even if intercepted.

While blockchain provides foundational security, BPM complements this by managing and controlling access to PII (see Fig. 3). BPM tools allow for the definition and enforcement of access policies, ensuring that only authorized users can view specific data. These tools can also monitor access patterns, detect anomalies, and flag any unauthorized access attempts. By integrating BPM with blockchain, pension systems can achieve a more robust level of data security and compliance.

For example, in a pension system where administrators, financial advisors, and auditors need varying levels of access to PII, the BPM engine can define access rights based on roles and responsibilities. All access requests and actions are recorded on the blockchain, providing a complete audit trail. If an unauthorized access attempt is detected, the BPM system can issue alerts and take corrective measures. This integration of blockchain and BPM ensures that PII is accessed solely by authorized individuals and only for legitimate purposes.

The integration of BPM with blockchain technology is visually represented in the following diagram Fig. 3:

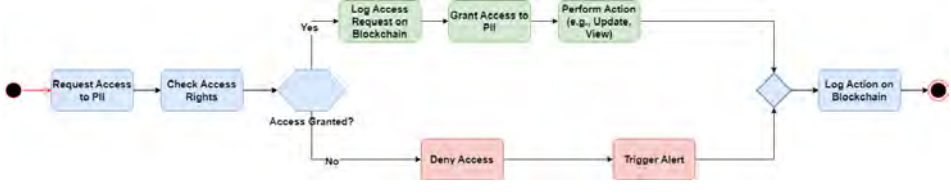


Figure 3: Data Access Management Workflow

6 Regulatory Compliance and Legal Considerations

Pension systems are required to adhere to strict regulatory frameworks designed to safeguard participants' interests and ensure the integrity of pension funds. These regulations

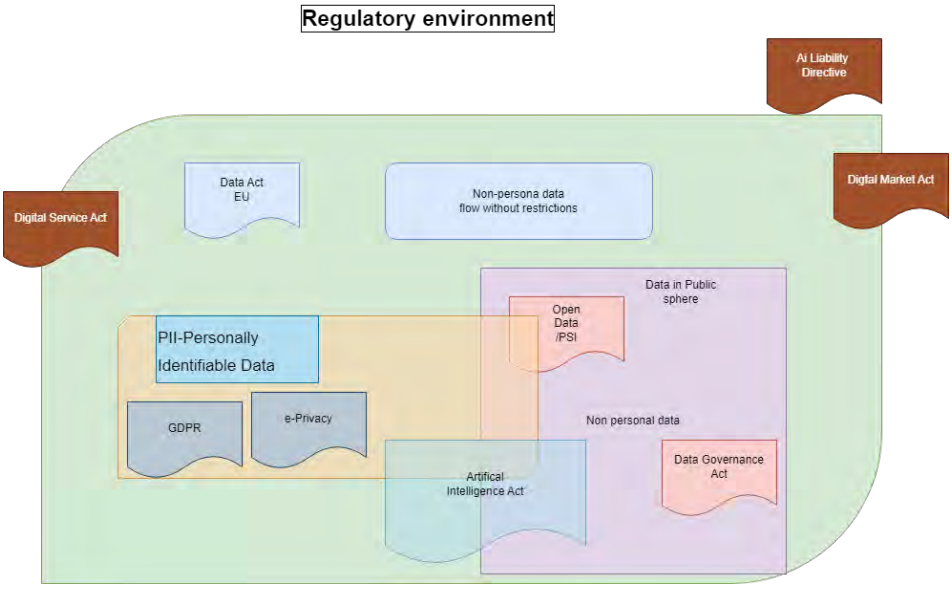


Figure 4: Regulatory Environment for Data Processing

typically mandate specific processes, reporting standards, and safeguards to protect Personally Identifiable Information (PII). The integration of BPM with blockchain technology offers a robust solution to meet these regulatory requirements while simultaneously enhancing the transparency and efficiency of pension systems.

Several regulatory frameworks govern the operations of pension systems, including the General Data Protection Regulation (GDPR), the California Consumer Privacy Act (CCPA), and the Employee Retirement Income Security Act (ERISA). BPM and blockchain technologies provide key mechanisms for ensuring compliance with these regulations [16, 25, 26].

The relevant regulations for data processing in the European Union that are incorporate into the local jurisdiction of the Member States.(a) **General Data Protection Act** [7], (b) **Open Data** [9] directive, (c) **Data Governance Act**[14] aims to facilitate data sharing across sectors, (d) **Data Act**[15] is more on business-to-consumer and business–to-business data sharing, (e) **Digital Markets Act (DMA)** [12] regulates the large digital platforms, (f) **AI Act** [10], (g) **AI Act** [10] is the risk mitigation of this disruptive technology, and (h) the directive on **AI Liability** [11] aims to define rules on the disclosure of evidence in the case of AI applications (see Fig. 4).

The legislative framework surrounding data privacy and regulation in the European Union has evolved significantly in recent years(see 4). Following the introduction of the General Data Protection Regulation (GDPR), other legislative initiatives have emerged to address various aspects of data privacy and management [16]. The ePrivacy Regulation aims to update the outdated ePrivacy Directive by complementing the GDPR in the realm of electronic communications. It regulates confidentiality, data storage, erasure, and permissible processing of electronic communication data[8]. The regulation on the free flow of non-personal data seeks to promote the free movement of non-personal data across the EU by establishing rules for data localization, availability to authorities, and data portability, distinguishing between personal and non-personal data. The Open

Data Directive replaces the PSI Directive and is designed to unlock the potential of public sector information, enabling real-time access to dynamic public data for reuse through advanced technical means[9]. The Data Governance Act creates a framework for sharing data across sectors, ensuring that trusted intermediaries facilitate this exchange while adhering to data protection regulations[14]. The Data Act focuses on business-to-consumer and business-to-business data sharing. It mandates that product and service-related data be accessible to users, granting them the right to share it with third parties[15]. The Digital Markets Act (DMA) regulates large digital platforms, such as search engines and cloud services, establishing fair competition rules, data accessibility, and interoperability obligations for platforms that serve as gateways between businesses and end-users[12]. The Digital Services Act (DSA), alongside the DMA, builds on the eCommerce directive and establishes a safer digital environment by harmonizing rules for intermediary service providers and introducing due diligence obligations for specific service providers[13]. In contrast to these technologically neutral regulations, the AI Act takes a more cautious approach due to concerns about the risks posed by artificial intelligence. The regulation reflects ongoing changes, even altering AI's legal definition during legislative discussions. It imposes significant administrative burdens on developers and providers. Complementing this, the AI Liability Directive establishes rules for evidence disclosure and the burden of proof in cases of damage caused by high-risk AI systems in civil claims. This complex legislative framework reflects the EU's efforts to balance innovation with regulation in the digital and AI spaces[10, 11]. BPM systems can automate the verification of all pension system operations to ensure they comply with relevant regulations. This includes ensuring that data handling practices meet GDPR and CCPA standards [30, 25]. Blockchain technology complements this by providing an immutable record of all transactions and actions taken within the pension system, ensuring both transparency and traceability [4]. BPM tools can also define and enforce access controls, ensuring that only authorized individuals have access to PII, and enforce data minimization principles to ensure that only necessary data is collected and processed [2].

For instance, in a pension system operating within the European Union, GDPR compliance is mandatory for protecting PII. By integrating BPM and blockchain, the pension system can achieve compliance through processes that ensure only the necessary data is collected, access to PII is tightly controlled, all actions involving PII are recorded on the blockchain, and comprehensive compliance reports are generated to demonstrate adherence to GDPR standards.

7 Challenges and Limitations

While the integration of BPM with blockchain technology in pension systems presents numerous advantages, it also comes with several challenges and limitations. Addressing these challenges is essential to formulating effective strategies that mitigate risks and ensure the smooth and successful implementation of these technologies [4].

One of the primary obstacles in combining BPM with blockchain technology lies in the technical complexity of the integration. Blockchain, being a relatively new and rapidly evolving technology, requires specialized expertise for its implementation, maintenance, and ongoing support. Achieving seamless interoperability between BPM tools, blockchain platforms, and existing legacy systems is another major technical hurdle that needs to be addressed [2].

Pension systems operate within stringent regulatory frameworks designed to protect

participants' interests and safeguard the integrity of pension funds. The integration of new technologies like BPM and blockchain must adhere to these regulations, which can be challenging due to the continually evolving nature of both technological innovations and regulatory requirements. Ensuring the privacy and security of PII is paramount, and strategies such as off-chain storage or the use of advanced encryption techniques are essential to meet these privacy requirements [16, 25].

Implementing BPM and blockchain technologies in pension systems also introduces significant operational changes, which can be difficult to manage effectively. These new technologies necessitate comprehensive change management strategies to ensure that all stakeholders—employees, pensioners, and regulators—are adequately informed and trained. Additionally, the initial cost of deploying BPM and blockchain solutions can be substantial, requiring organizations to carefully assess the cost-benefit ratio and allocate resources effectively to support the rollout [2].

Though blockchain technology offers enhanced security features, it is not completely immune to risks. Blockchain networks can still be vulnerable to certain types of cyberattacks, such as 51% attacks, and smart contracts can potentially contain vulnerabilities that attackers may exploit [18, 22]. Therefore, implementing robust cybersecurity measures and ensuring the secure development and rigorous auditing of smart contracts is crucial to maintaining the system's integrity [4].

Despite these challenges, ongoing research and development efforts are focused on overcoming these limitations and enhancing the integration of BPM and blockchain in pension systems. Current research initiatives are working on scalable blockchain solutions, providing clearer regulatory guidance, and implementing more advanced security measures to improve the overall reliability and efficiency of these systems [4].

8 Summary and Conclusion

The integration of BPM with blockchain technology in pension systems enhances transparency, efficiency, and compliance. However, to fully realize these benefits, organizations must address a variety of technical complexities, regulatory uncertainties, and operational challenges. By leveraging the combined strengths of these two technologies, pension systems can be restructured to serve stakeholders more effectively, while operating with greater efficiency and transparency [4, 2].

A comprehensive planning approach is essential for the successful integration of BPM and blockchain, with attention given to technical, regulatory, and operational factors.

The successful implementation of these technologies hinges on the engagement and support of all stakeholders, including pensioners, administrators, and regulators. Continuous monitoring and ongoing process optimization are crucial to adapt to the ever-evolving regulatory environment and technological advancements. Additionally, organizations should prioritize investing in expertise in both BPM and blockchain technologies, ensuring that they are prepared for effective implementation and long-term maintenance. Robust security measures must also be established to protect sensitive data and maintain the integrity of the system [16, 25].

In conclusion, while the integration of BPM and blockchain in pension systems is still in its early stages, the potential advantages make it a promising field. By addressing the inherent challenges and capitalizing on the strengths of these technologies, pension systems can be transformed into more efficient and transparent operations that better serve their stakeholders [4, 2].

8.1 Scientific contribution

The project objective was to carry out a software experiment in the form of industrial research and development. The research question was whether the available technologies make it possible to create an IT System that can support a complex business process. The development and implementation reduced the uncertainty of the feasibility of the solution. Based on Business Process Management and Modeling, an Enterprise and IT architecture was defined. Service Oriented Computing was utilized to reconcile the business process requirements and the services of the underlying blockchain technology [28]. Our model is based on the Business Process Modeling and the representation of the processes by Business Process Modelling Notation [31]. We have selected this solution since the representation of the processes makes it possible to do model checking [23]. We combined the workflow and blockchain patterns in a Service Oriented, Web service (API) based framework [1, 32, 6]. At the time of the research and development, the comprehensive auditing and validation of decentralized applications was not available. This is a limitation of the research and denotes future work in the domain of cybersecurity, reliability, and trustworthiness of the solution.

The main goal of the project was to conduct a software experiment through industrial research and development. The research sought to determine whether existing technologies could be used to create an IT system capable of supporting a complex business process. The development and implementation phases helped to reduce uncertainty around the feasibility of the solution. Building upon Business Process Management and Modeling, we defined an enterprise and IT architecture. Service Oriented Computing was employed to align business process requirements with the services provided by the underlying blockchain technology [28].

Our model is based on Business Process Modeling and the representation of the processes through Business Process Modelling Notation [31], which enables model checking [23]. We chose this solution because the representation of the processes allows for model checking. We integrated workflow and blockchain patterns in a Service Oriented, web service (API) based framework [1, 32, 6].

At the time of the research and development, a comprehensive auditing and validation process for decentralized applications was not available. This limitation of the research identifies future work in the domain of cybersecurity, reliability, and trustworthiness of the solution [5]. Our research demonstrates the plausibility of the application of blockchain and smart contract technologies with the integration of tools in Enterprise Architecture frame.

9 Data Availability Statement:

The data and code presented in this study are available upon request from the authors [33].

10 Acknowledgments

This paper was supported by the "Application Domain-Specific Highly Reliable IT Solutions" project, implemented with the assistance of the National Research, Development and Innovation Fund of Hungary. The project was financed under the Thematic Excellence Programme TKP2021-NVA-29 (National Challenges Subprogramme) funding

scheme. Additional support was provided by NKFIH-KFI-2018.1.1.2-KFI-2018-0056 and by the COST Action CA19130.

References

[1] Wil Aalst. *Workflow management : models, methods, and systems*. MIT Press, Cambridge, Mass, 2002.

[2] Digital Adoption. A guide to business process management tools. `https://www.digital-adoption.com/business-process-management-tools/`, 2021.

[3] Shuzhi Cheng, Wenxuan Shi, and Hualong Zhang. Voltimebank: A volunteer system for mutual pension based on blockchain. In *Proceedings of the 2019 International Conference on Blockchain Technology*, ICBCT 2019. ACM, March 2019.

[4] Pierre Edrud. Improving bpm with blockchain technology: Benefits, costs, criteria & barriers. Technical report, Mid Sweden University, 2021.

[5] Enterprise Ethereum Alliance Inc. EEA EthTrust Security Levels Specification version 2, December 2023. [Online; accessed 17. Aug. 2024].

[6] Thomas Erl. *Service-oriented architecture : analysis and design for services and microservices*. Prentice Hall, Service Tech Press, Boston, 2017.

[7] European-Parliament "and" of the Council. Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/ec (general data protection regulation). *Regulation (eu)*, 679:2016, 2016. application 25 May 2018.

[8] European-Parliament "and" of the Council. Proposal for a Regulation of the European Parliament and of the Council concerning the respect for private life and the protection of personal data in electronic communications and repealing Directive 2002/58/EC (Regulation on Privacy and Electronic Communications), 2017. application = TBD, [Online; accessed 20. Feb. 2024].

[9] European-Parliament "and" of the Council. Directive - 2019/1024 - EN - psi directive - EUR-LexDirective (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information, February 2019. application 17 July 2021[Online; accessed 20. Feb. 2024].

[10] European-Parliament "and" of the Council. Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts, 2021. application = TBD, [Online; accessed 20. Feb. 2024].

[11] European-Parliament "and" of the Council. Proposal for a Directive of the European Parliament and of the Council on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive), 2022. application = TBD, [Online; accessed 20. Feb. 2024].

[12] European-Parliament "and" of the Council. Regulation - 2022/1925 - EN - EUR-Lex Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act), February 2022. application 2 May 2023, [Online; accessed 20. Feb. 2024].

[13] European-Parliament "and" of the Council. Regulation - 2022/2065 - EN - DSA - EUR-LexRegulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act), 2022. application 2 February 2024, [Online; accessed 20. Feb. 2024].

[14] European-Parliament "and" of the Council. Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act), 2022. application 24 September 2023, [Online; accessed 20. Feb. 2024].

[15] European-Parliament "and" of the Council. Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 on harmonised rules on fair access to and use of data and amending Regulation (EU) 2017/2394 and Directive (EU) 2020/1828 (Data Act), 2023. application 12 September 2025, [Online; accessed 20. Feb. 2024].

[16] European Union. General Data Protection Regulation (GDPR). Technical report, European Union, 2018.

[17] Dongchen Gong and Gongrang Zhang. Research on the application value of blockchain to for-profit pension institutions: Focus on blockchain's impacts on service mode upgrading process. In *Proceedings of the Fifth International Conference on Economic and Business Management (FEBM 2020)*, febm-20. Atlantis Press, 2020.

[18] Yuechen Hao. Research of the 51In *2022 3rd International Conference on Computer Vision, Image and Deep Learning ; International Conference on Computer Engineering and Applications (CVIDL &; ICCEA)*, volume 31, pages 278–283. IEEE, May 2022.

[19] Alan Hevner and Samir Chatterjee. *Design Research in Information Systems*. Springer US, 2010.

[20] Rui Hu, Xinyi Liang, Sen Wang, Yu Liu, and Jin Dong. Design and implementation of access control for pension data: A smart contract approach. In *2023 8th IEEE International Conference on Network Intelligence and Digital Content (IC-NIDC)*. IEEE, November 2023.

[21] Ridong Huang, Jianmao Xiao, Siqi Cheng, Ronglin Zhang, Hao Zeng, Xin Huang, and Zuoyi Liao. *A New Home-Based Pension Mutual Aid Framework Based on Blockchain*, pages 32–47. Springer Nature Switzerland, 2023.

[22] Naghma Khatoon, Rishav Rishu, Sushant Verma, and Prashant Pranav. Proposing a modified proof of stake system to counter 51 In *2024 2nd International Conference on Device Intelligence, Computing and Communication Technologies (DICCT)*, pages 01–04. IEEE, March 2024.

[23] Meriem Kherbouche and Bálint Molnár. Formal model checking and transformations of models represented in uml with alloy. In *Modelling to Program: Second International Workshop, M2P 2020, Lappeenranta, Finland, March 10–12, 2020, Revised Selected Papers 1*, pages 127–136. Springer, 2021.

[24] Bálint Molnár, Galena Pisoni, Meriem Kherbouche, and Yossra Zghal. Blockchain-based business process management (bpm) for finance: the case of credit and claim requests. *MDPI Smart Cities*, 6:1254–1278, 2023.

[25] State of California Department of Justice. California consumer privacy act (ccpa). Technical report, State of California Department of Justice, 2020.

[26] U.S. Department of Labor. Employee retirement income security act (erisa). Technical report, U.S. Department of Labor, 1974.

[27] Indranil Sarker and Bidisha Datta. Re-designing the pension business processes for achieving technology-driven reforms through blockchain adoption: A proposed architecture. *Technological Forecasting and Social Change*, 174:121059, January 2022.

[28] Munindar P Singh and Michael N Huhns. *Service-oriented computing: semantics, processes, agents*. John Wiley & Sons, Chichester, England, 2005. Use copy Restrictions unspecified star MiAaHDL.

[29] Wil M.P. Van Der Aalst, Marcello La Rosa, and Flávia M. Santoro. Business process management. *Business & Information Systems Engineering*, 58(1):1–6, 2016.

[30] Paul Voigt and Axel von dem Bussche. *The EU General Data Protection Regulation (GDPR)*. Springer International Publishing, 2017.

[31] Stephen A. White and Derek Miers. *BPMN modeling and reference guide*. Future Strategies Inc., Lighthouse Point, Fla., 2008.

[32] Wei Xu. Toward human-centered ai: a perspective from human-computer interaction. *interactions*, 26(4):42–46, 2019.

[33] Ábris Informatikai és Tanácsadó Korlátolt Felelőségű Társaság. Projekt nkfih-kfi-2018.1.1.2-kfi-2018-0056 a blokklánc technológiai megoldások vizsgálata és fejlesztése nyugdíjbiztosítási információk és a tudományos publikációk tárolásának területén (Testing and developing blockchain technology solutions for the storage of pension information and scientific publications), September 2019. Accessed:2021-02-12.

Can social exclusion be revealed from studying brain activity via EEG?

Mateja Lesar[1], Teja Štrempfel[1], Larisa Moderc[2], Lynn Süthoff[2], Jakob Sajovic[3], Peter Rogelj[4], Gorazd Drevenšek[5], Zoran Levnajić[1]

[1]Complex Systems and Data Science Lab, Faculty of Information Studies
Ljubljanska cesta 31a,8000 Novo mesto, Slovenija
{mateja.lesar, teja.strempfel, zoran.levnajic}@fis.unm.si

[2]University of Vienna, Austria
Universitätsring 1, 1010 Wien
larisa.moderc@gmail.com, a01624726@unet.univie.ac.at

[3]Department of Orthodontics, University Medical Centre Ljubljana,
Hrvatski trg 6, Ljubljana, Slovenija
jakob.sajovic@kclj.si

[4]Faculty of Mathematics, Natural Sciences and Information Technologies, University of Primorska
Glagoljaška ulica 8, 6000 Koper, Slovenija
peter.rogelj@upr.si

[5]The Institute of Pharmacology and Experimental Toxicology, Faculty of Medicine, University of Ljubljana,
Korytkova cesta 2, 1000 Ljubljana, Slovenija
gorazd.drevensek@mf.uni-lj.si

Abstract: *Social exclusion is an important issue in modern society, but its neurological underpinnings remain unclear. We explored whether social exclusion can be detected in EEG data collected from participants engaged in a Cyberball game – a gaming framework designed to simulate social exclusion in unsuspecting participants. We transformed the EEG data into functional connectivity networks and applied machine learning to identify connectivity patterns potentially linked to the experience of social exclusion. In this paper, we present our preliminary findings, which suggest that such patterns can indeed be recognized. We anticipate further confirmation and extension of these results in our future work.*

Key Words*: social exclusion, Cyberball game, EEG measurements, functional connectivity networks, machine learning*

1 Introduction

Social exclusion is common in everyday life. It occurs in human and nonhuman animals [1,2] and exists in every culture [3]. Nezlek et al. reported [4] that people are ostracized daily – typically that means that we all have one experience of rejection or disregard per day. Being socially excluded breaks the relationships we have and could influence our physical and mental health. Social exclusion increases depression, causes aggression and impulsive behavior, decreases survival rates, and possibly promotes suicides [5]. In addition, socially excluded people are less self-regulated and show a decrease in the mental processes involved in cognitive control [6].

An electroencephalogram (EEG) measures the electrical activity of the brain via electrodes placed on the scalp. Beta frequencies, which range from 13 to 30 Hz, are linked to active cognitive processes such as focused attention, problem-solving, and heightened alertness. These brainwave patterns are believed to be particularly relevant in social exclusion situations. We hypothesize that social exclusion could be identified by studying these patterns.

In our experiment, we recorded EEG data from 32 participants while they played a *Cyberball* game, a widely used experimental game in social psychology and neuroscience to simulate social exclusion [7]. In *Cyberball*, participants think they are playing a virtual ball-tossing game with other online players, but the players are in fact computer-generated [8]. Participants either experience equal inclusion in the game or are excluded from a majority of the ball tosses. Our study involved both versions of this game, with EEG recorded throughout each session to examine brainwave patterns during inclusion and exclusion conditions.

2 Cyberball game and EEG measurements

2.1 Participants

The sample included 32 right-handed participants (22 females, 10 males, age range: 20–40, mean age = 24.9, SD = 5.1). All participants met the inclusion and exclusion criteria (see authors for details).

2.2 Cyberball

In the Cyberball game, the researcher controls the number of throws using the Cyberball GUI (Cyberball 5 Manual, 2019). Participants playing the Cyberball game were exposed to 204 ball throws receiving 33% of ball passes in the inclusion condition and 17% in the exclusion condition. The modification to our experimental paradigm was programmed at the HICUP lab, Faculty of Mathematics, Natural Sciences, and Information Technologies, Koper, Slovenia.

2.3 Brain Activity Measurements

Brain activity was recorded using a 32-electrode EEG cap (g.NAUTILUS system) while participants played a 15-minute Cyberball game. In the first game, participants experienced the social inclusion condition, while in the second, they were subjected to the exclusion condition. During each game, participants pressed a button to pass the ball to one of the two computer-generated players. For specific details on the EEG setup, electrode placement, and equipment, refer to the authors.

Preprocessing for artifact and noise removal followed standard protocols using the EEGLAB toolbox in MATLAB. The time-series data was segmented into intervals from 200 ms before to 1000 ms after a computer player caught or passed the ball (focusing only on throws between the computer players where the participant was neglected). EEG data was filtered into the following frequency bands: theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz) low gamma (30–45 Hz), as well as the full spectrum (1-50 Hz). For the purposes of this short paper, we look only at beta bandpass.

3 Brain connectivity

We calculated functional connectivity across 32 time-series of all epochs. This short paper focuses solely on the Complex Pearson Correlation Coefficient measure of connectivity (CPCC) in [9]. CPCC provides complex-valued results, where absCPCC corresponds to brain processes that are either connected and anatomically close or similar and occurring simultaneously. This measure is however vulnerable to confounding by volume conduction. imagCPCC on the other hand represents brain processes that are distant, more likely reflecting information transfer through white matter fibers and unaffected by volume conduction. This allows for a nuanced analysis of connectivity, distinguishing between real and imaginary components [9, 10]. We present preliminary results for the connectivity network obtained via the CPCC.

Each electrode has time series data. We computed connectivity CPCC measure for each electrode pair giving us 32x32x512/2 connectivity matrices. For each epoch, we have data to which condition it belongs. For each epoch, we calculated the connectivity matrix, which was used as an input layer in ML described below in Chapter 4.

4 Analyzing networks via Machine learning

Machine learning (ML) was conducted in MATLAB aimed to classify the EEG data based on the inclusion and exclusion conditions using MATLAB's Deep Learning Toolbox (The MathWorks, Natick, MA, USA. As an input feature connectivity matrices for the CPCC beta bandpass are selected. The architecture of our machine learning model was a 3-layer fully connected convolutional neural network (CNN), with the first (input) layer receiving connectivity measures as input features. The final output layer classified each epoch into one of two conditions: exclusion or inclusion. The goal was to classify whether the EEG data belonged to the inclusion or exclusion condition of the Cyberball game the participants were engaged in. That is to say, we wish to check if we can train an ML algorithm to recognize and differentiate between social inclusion and exclusion only from this data.

The dataset was split into training (80%) and testing (20%) sets. To optimize model performance, we used the stochastic gradient descent with momentum (SGDM) optimizer. The maximum number of training epochs was set to 80, with data shuffling at each epoch. We performed five repetitions assessing variability using five-fold cross-validation. We did not use the third-validation subset, as we did not optimize the classifier metaparameters.

5 Preliminary results and discussion

We present our results of machine learning. The machine learning algorithm showed great performance on connectivity features from the CPCC method, that is, 98.60 %. Fig. 1 represents training progress.

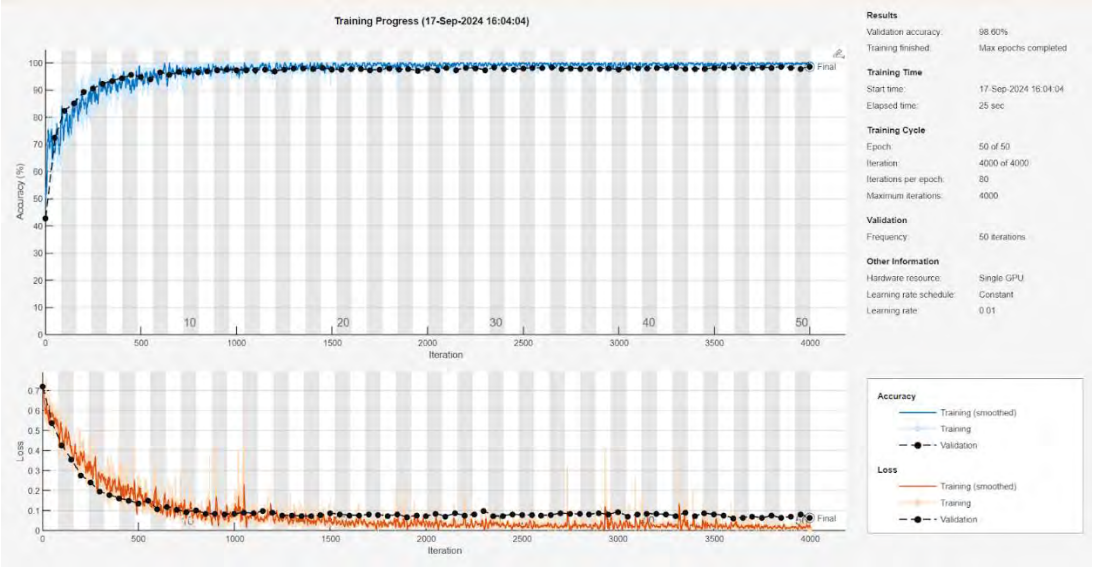


Figure 1. The training progress utilizes the CPCC method, achieving a validation accuracy of 98.90%. The training cycle comprised 50 epochs with 4,000 iterations, averaging 80 iterations per epoch. Validation was conducted every 50 iterations, using a single CPU for hardware resources, and the learning rate was maintained at a constant value of 0.01.

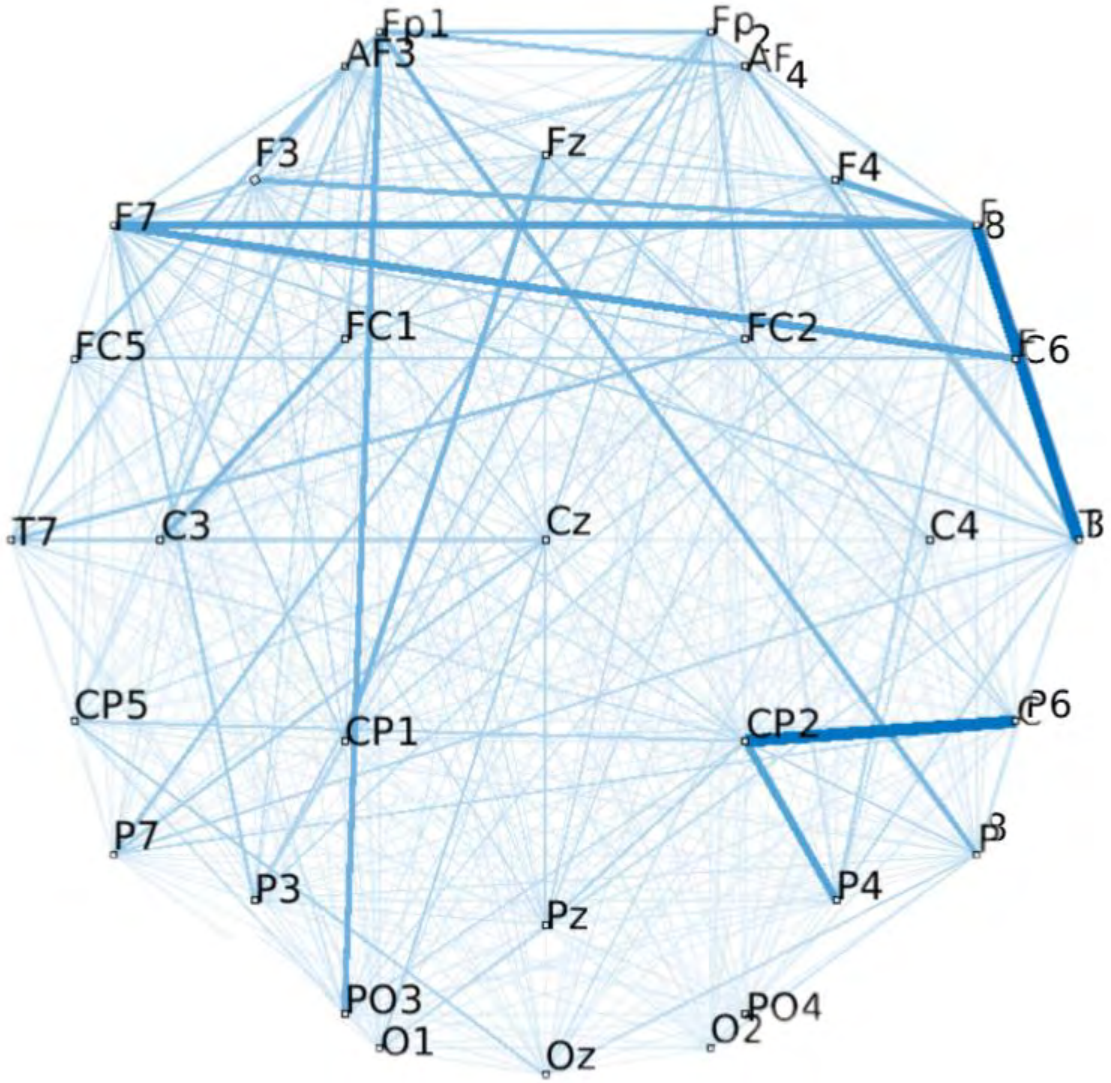


Figure 2: Feature importance of the CPCC connectivity matrix in the beta frequency band. The 32 electrodes are placed according to the 10/20 system, starting in the frontal lobe and ending in the occipital lobe (F stands for frontal, C for central, P for parietal, and O for occipital electrodes.

Fig. 2 illustrates the CPCC connectivity features identified by the neural network algorithm. The thickness of the lines represents the importance of connectivity between different regions. Notably, the strongest connections are observed between electrodes F8 and T8, Cp2 and CP6, F7 and F8, F7 and FC6.

Preliminary results from the beta frequency band show consistent connectivity between key electrodes, particularly F8 and T8, which are linked to social cognition and emotion regulation within the social pain network [5]. Connections involving F7, FC6, and parietal electrodes CP2 and CP6 also suggest involvement in emotion regulation and social pain processing [11]. These findings highlight the role of brain regions associated with both social and physical pain in the experience of social exclusion [11,12].

These connectivity findings provide insights into the neural mechanisms underlying social exclusion, particularly in regions involved in social cognition and emotion regulation. Our results show that the ML algorithm could indeed differentiate between inclusion and exclusion conditions. The study has certain limitations and careful interpretation is needed with the primary limitation being a relatively small sample size. First, while we plan to address this by balancing inclusion and exclusion conditions using duplication or bootstrapping techniques, a

larger dataset could enhance generalizability. We already utilized subject-variant and subject-invariant approaches to address model validation. Additionally, ML using multiple connectivity measures and frequency bands was already performed to compare the results of the model's classification. Saliency maps provided insight into neural differences between inclusion and exclusion in the 0.2–0.6 ms window, aligning with the N200 and P300 event-related potentials (ERPs) reported in other Cyberball studies [5]. With resting state data, a baseline condition of saliency networks could be identified for each participant. This data could serve as a baseline isolating socially driven neural effects and will be used in our future analyses. The machine learning methodology may provide valuable insight into a complex system of brain networks involved in neural mechanisms of social exclusion highlighting its potential applications in mental health interventions (such as targeted personalized intervention, early detection, neurofeedback etc.). A forthcoming, comprehensive scientific paper will present all future results and a detailed discussion of the study's limitations.

6 Acknowledgements

This work received financial support by the Slovenian Research Agency (ARRS) via Program P1-0383 and P3-0293(B), and via Project J5-8236. We are especially grateful to all participants who took part in our experiment. We thank the HICUP lab at the Faculty of Mathematics, Natural Sciences, and Information Technologies, Koper, Slovenia, for their technical support in customizing the experimental paradigm.

7 References

[1] Kipling D. Williams, Chapter 6 Ostracism: A Temporal Need-Threat Model, Advances in Experimental Social Psychology, Academic Press, Volume 41, 2009, Pages 275-314.
[2] Williams, K. D., & Nida, S. A. (2011). Ostracism: Consequences and coping. Current Directions in Psychological Science, 20(2), 71–75.
[3] Gruter, M., & Masters, R. D. (1986). Ostracism as a social and biological phenomenon: An introduction. Ethology & Sociobiology, 7(3-4), 149–158.
[4] Nezlek, John B., Wesselmann, Eric D., Wheeler, Ladd, Williams, Kipling D. Ostracism in everyday life. Group Dynamics: Theory, Research, and Practice, Vol 16(2), Jun 2012, 91-104.
[5] Kawamoto, T., Ura, M., & Nittono, H. (2015). Intrapersonal and interpersonal processes of social exclusion. Frontiers in neuroscience, 9, 62.
[6] Otten, M., & Jonas, K. J. (2013). Out of the group, out of control? The brain responds to social exclusion with changes in cognitive control. Social Cognitive and Affective Neuroscience, 8(7), 789–794.
[7] Williams, K. D., & Jarvis, B. (2006). Cyberball: A program for use in research on interpersonal ostracism and acceptance. Behavior Research Methods, 38(1), 174–180.
[8] Wesselmann, E. D., Williams, K. D., & Hales, A. H. (2013). Vicarious ostracism. Frontiers in Human Neuroscience, 7, 153.
[9] Šverko Z, Vrankić M, Vlahinić S, Rogelj P. Complex Pearson Correlation Coefficient for EEG Connectivity Analysis. Sensors. 2022; 22(4):1477.
[10] Močilnik, V., Rutar Gorišek, V., Sajovic, J., Pretnar Oblak, J., Drevenšek, G., & Rogelj, P. (2024). Integrating EEG and Machine Learning to Analyze Brain Changes during the Rehabilitation of Broca's Aphasia. Sensors (Basel, Switzerland), 24(2), 329.
[11] Eisenberger, N. The pain of social disconnection: examining the shared neural underpinnings of physical and social pain. Nat Rev Neurosci 13, 421–434 (2012).
[12] Eisenberger N. I. (2015). Social pain and the brain: controversies, questions, and where to go from here. Annual review of psychology, 66, 601–629.

Assessing Pigmented Skin Moles Using Minkowski Fractal Dimensions and Comprehensive Color Channel Analysis

Mia Darkovska, Ilija Mizhimakoski
Faculty of Electrical Engineering and Information Technologies,
Ss. Cyril and Methhodius Univ. in Skopje
Rugjer Boshkovikj 18, 1000 Skopje, North Macedonia
darkovskam@gmail.com, ilijamizhimakoski@gmail.com
Ana Ristovska Dimitrovska
Faculty of Medical Sciences, Ss. Cyril and Methhodius Univ. in Skopje
General City Hospital 8 mi Septemvri
Bledski dogovor, 1000 Skopje, North Macedonia
nanaristovska@yahoo.com
Jasmina Angelevska Kostadinoska, Vesna Andova
Faculty of Electrical Engineering and Information Technologies,
Ss. Cyril and Methhodius Univ. in Skopje
Rugjer Boshkovikj 18, 1000 Skopje, North Macedonia
{jasminaa, vesnaa}@feit.ukim.edu.mk

Abstract. *Distinguishing between benign and malignant moles is crucial for early detection and treatment of skin cancer, as malignant moles can develop into life-threatening conditions if not identified promptly. Doctors often use the ABCDE criteria (Asymmetry, Border, Color, Diameter, and Evolution) to assess and differentiate these moles. In the analysis of skin tumors, the C criterion (color) plays a critical role in distinguishing between benign and malignant moles. This study examines the effectiveness of various color spaces, such as HSV, XYZ, YCbCr, and Lab, in the differentiation process. Our research indicates a statistically significant difference in the channel features of these color spaces when comparing benign to malignant skin moles. By evaluating these differences, we demonstrate that color can serve as a reliable factor in identifying the two types of moles. Additionally, we will briefly discuss the B criterion (border irregularity) to provide a broader perspective on mole classification.*

Keywords. melanoma, image analysis, color spaces, Minkowski dimension, color channels

1 Introduction

Skin cancer, characterized by abnormal skin cell growth, is primarily linked to sun exposure but can also develop in sun-protected areas. Malignant skin tumors are infiltrative, invasive, and poorly demarcated, with potential to metastasize, invade neighboring tissues, and recur. They consist of atypical, anaplastic cells and are classified as pigmented (e.g., melanoma malignum) or non-pigmented (e.g., basal and squamous cell carcinomas).

Melanoma malignum, one of the most aggressive skin cancers, originates from melanocytes and has high metastatic potential. In 2020, 325,000 melanoma cases and 57,000 deaths were reported globally [16], with rising incidence among older adults and young populations alike [6]. De Pinto et al. [6] observed stable or increasing incidence across most countries, with mortality declining by $40-50\%$ in Australia, highlighting that early diagnosis and prevention can significantly improve outcomes.

Given the aggressive nature of melanoma, early detection is vital. Tumor classification often utilizes the ABCDE criteria [8]: "A" for asymmetry, as malignant melanomas tend to lack symmetry; "B" for border irregularity, where uneven or notched edges are common; "C" for color, as multiple shades or colors in a lesion raise suspicion; "D" for diameter, with melanomas typically larger than 6 mm; and "E" for evolution, indicating changes in size, shape, or color over time.

Computer-aided diagnosis systems can be very helpful for early melanoma detection. Numerous algorithms have been proposed to distinguish melanoma from benign pigmented moles in dermoscopic images, but they face challenges when analyzing early-stage lesions. Some studies focus on the "B" criteria, or border irregularity, using fractal dimensions [5, 13, 12], including the Minkowski dimension of the mole or its border. This method has shown that malignant moles have significantly different fractal dimensions than benign ones [5], highlighting its diagnostic potential. However, color analysis is equally vital for melanoma detection. In [3], a method based on mole color and texture was proposed, using automatic segmentation via k-means clustering. The study demonstrated superior performance over earlier approaches [1, 2, 9]. Recent advances involve machine learning and deep learning methods [12, 13, 15].

This study builds on previous work by combining segmentation techniques and binarization methods to isolate moles from surrounding skin for further analysis. In addition to assessing the Minkowski dimension for geometric complexity, we emphasize color variation. Malignant moles typically show more diverse pigmentation than benign ones, making color models such as HSV and YCbCr valuable for edge detection and malignancy differentiation. This enhanced color analysis is critical for improving early diagnosis of melanoma. The purpose of this study is not to develop a machine learning model, but rather to employ statistical analysis of Minkowski dimensions and color features to improve the detection of malignant versus benign moles. The paper is organized as follows. In Section 2-Preliminaries, we examine various color models and their suitability for analyzing mole pigmentation, along with a brief introduction to the Minkowski dimension. Section 3, Dataset and Methodology, details the dataset and the methods used in our analysis. In Section 4, Results, we present the hypothesis tests and discuss the findings. Finally, Section 5, Discussion and Conclusion, provides a comprehensive discussion of the results and concludes the study.

2 Preliminaries

Color spaces: Color spaces are mathematical models that systematically represent and manipulate colors as coordinates in a defined multidimensional space, facilitating the organization and communication of color information. Each color space encompasses a specific gamut and is defined by values in three dimensions or more. The applications of color spaces include image processing for tasks like edge detection, color manipulation for precise adjustments, and color management to ensure consistency across devices. The *HSV* (Hue, Saturation, Value) color model provides an intuitive way to represent and manipulate colors, where *Hue* indicates the color type, *Value* measures luminance, and *Saturation* defines color purity or vividness. Saturation refers to the intensity or purity of a color, indicating how much gray is present in it. A color with high saturation appears vivid and rich, while a color with low saturation appears more washed out or muted. *YCbCr* separates an image into luminance (Y) and chrominance (Cr, Cb) components, with Y encoding brightness and Cr/Cb representing color differences. This separation is beneficial for edge detection, as the Y channel contains most structural information. The *Lab* color model approximates human vision by splitting color into three components: L (luminance), a (green-red axis), and b (blue-yellow axis), making it effective for edge detection like YCbCr. The *XYZ* color model is a linear, device-independent space serving as the foundation for many other color models, encompassing the entire visible spectrum and ensuring precise color reproduction in applications such as high-end printing and imaging. A study [4] demonstrated that the percent melanoma color and color clustering ratio features yield comparable discrimination results, with the percent melanoma color feature achieving discrimination rates of 89% and 88% in the boundary areas of lesions. These findings suggest that variations in color saturation, particularly in the boundary regions, provide critical insights for melanoma discrimination and may enhance the understanding of color uniformity in lesion characterization. In the context of skin lesions, saturation can provide valuable information about the overall appearance of the mole. Malignant moles often display a greater range of pigments and colors, making them more visually striking. This variability can be an effective feature for distinguishing malignant lesions from benign ones, which typically have a more uniform coloration. The findings by Pillay and Viriri [14] underscore this notion, as malignant lesions are characterized by their multicoloration, further validating the significance of saturation in melanoma detection. They employed a scoring system based on the presence of distinct colors in the segmented image, where each color must meet specific threshold values. This approach, however, encounters challenges due to the non-linear and discontinuous nature of the RGB color space, making hue changes difficult to track and susceptible to illumination variations.

Minkowski dimension: The Minkowski dimension, or box-counting dimension, defined by Minkowski and Bouligand [10], is useful for analyzing pigmented skin moles as it quantifies roughness, a trait linked to malignant tumors [5]. If it exists for a set S, it can be calculated using:

$$D = \lim_{\varepsilon \to 0} \frac{\log N(\varepsilon)}{\log\left(\frac{1}{\varepsilon}\right)},$$

where $N(\varepsilon)$ is the minimum number of boxes of side length ε needed to cover S. A higher dimension indicates greater roughness and complexity [10]. This study expands on [5], focusing on the shape and roughness of skin moles to differentiate malignant from benign tumors.

Mole images are approximations due to resolution limits. Scale-dependent variations in real-world structures have been noted since the inception of fractal geometry. Mandelbrot [10] suggested that strict scale invariance holds only within specific bounds shaped by processes. Thus, careful preprocessing is vital for accurately approximating the Minkowski dimension.

3 Dataset and Methodology

The dataset employed in this study consists of 100 dermoscopic images of skin moles, obtained using a FotoFinder dermatoscope [7] at the "8-mi Septemvri" City Hospital in Skopje, North Macedonia. All images are captured by positioning the lens directly against the skin. Of the total dataset, 62 moles were clinically diagnosed as benign, and 38 were identified as malignant by a certified dermatologist. Prior to analysis, each image underwent preprocessing to isolate the mole from the surrounding skin and any extraneous elements. This preprocessing step is critical for ensuring accurate border segmentation and color analysis, which are key to assessing the morphological characteristics and potential malignancy of the moles.

In the previous work [5], we utilized grayscale conversion to segment the mole from the background. However, in this approach, we aimed to improve segmentation accuracy by converting the images to an alternative color space, such as HSV, YCbCr, Lab, and XYZ. The luminance channel from all of the mentioned color spaces proved valuable for mole-background segmentation, which enabled more accurate delineation of the mole's edges.

The following steps, along with the associated tools, were undertaken:

1. The relevant luminance or intensity channel was extracted from the converted color space. This step isolates the lightness information of the image, which is essential for the following steps.

2. To reduce noise in the luminance channel, we applied Gaussian blur filter. Specifically, the 'GaussianBlur' function from the OpenCV library in Python, with a kernel size of 5×5 and a standard deviation of 0. These parameters were determined through empirical testing, as they provided optimal noise reduction for this dataset without excessive blurring.

3. The denoised luminance channel was then converted into a binary image using Otsu's thresholding method. Although Contrast Limited Adaptive Histogram Equalization (CLAHE) was considered to enhance contrast prior to thresholding, it did not yield improved results for this particular dataset and was thus omitted in favor of direct thresholding.

4. Contours were extracted from the thresholded image using the `cv2.findContours` function. The largest contour was assumed to correspond to the mole's boundary. While alternative edge detection techniques, such as the Canny edge detector, were explored, `findContours` provided more reliable and consistent results.

5. A binary mask was created to isolate the mole region within the image based on the largest detected contour.

6. The segmented mole was cropped from the original image and resized to a standard dimension of 300×300 pixels. This ensures uniformity in size for further analysis and feature extraction steps.

All of the steps were executed using Python code [11]. HSV and XYZ segmentation of one benign mole is illustrated in Figures [1, 2], while Lab and YCrCb segmentation of one malignant mole is illustrated in Figures [3,4].

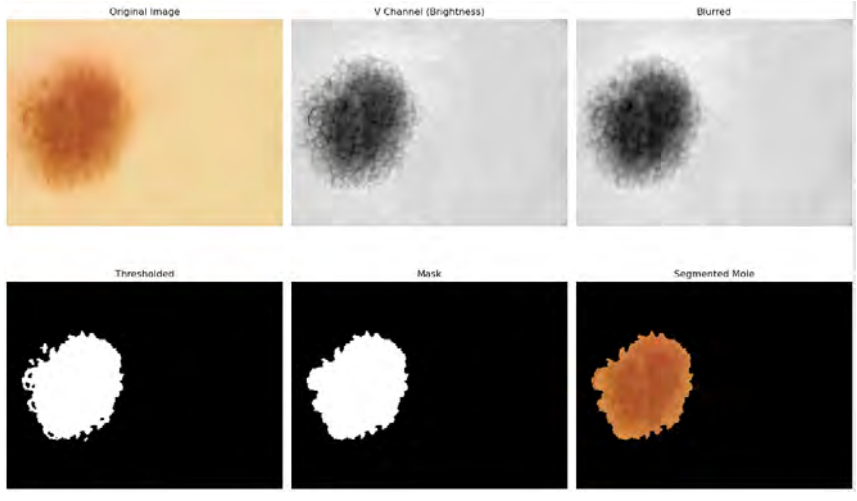


Figure 1: HSV Segmentation (Benign)

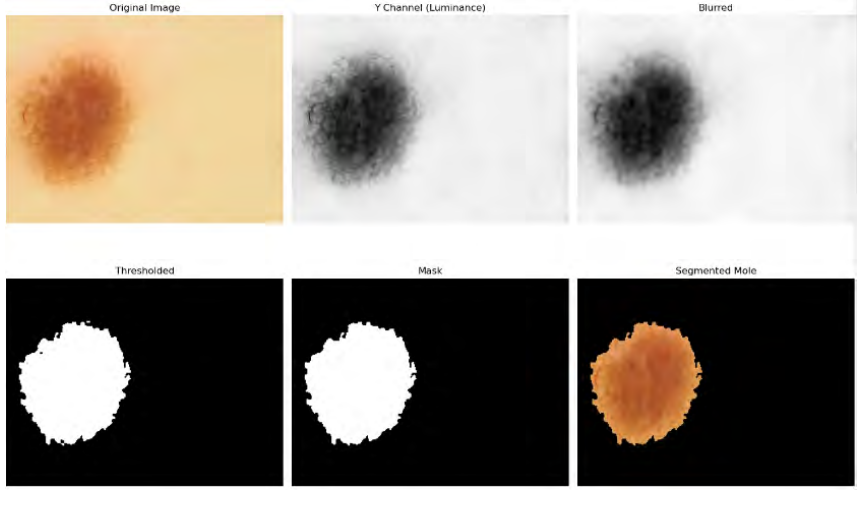


Figure 2: XYZ Segmentation (Benign)

It is important to note that certain preprocessing steps, such as noise reduction and edge extraction, may have introduced minor variations in the bounding Minkowski dimensions. However, since all images were subjected to the same standardized processing

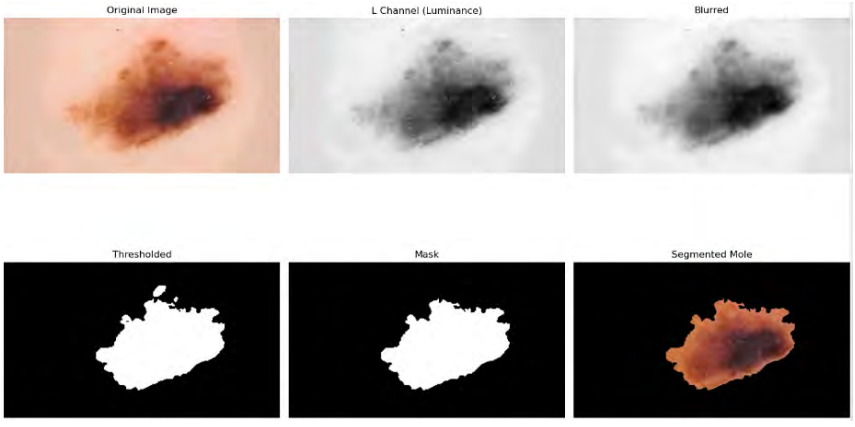


Figure 3: Lab Segmentation (Malignant)

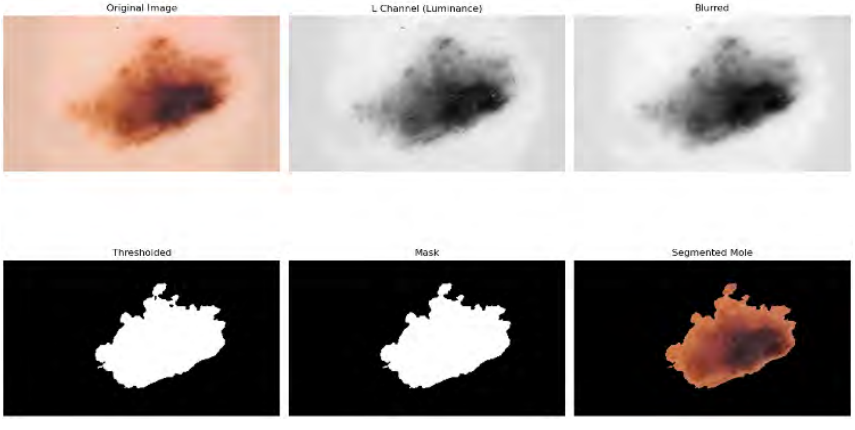


Figure 4: YCrCb Segmentation (Malignant)

pipeline, these variations are considered consistent across the dataset, and their impact on subsequent analyses is deemed negligible. Following the preprocessing steps, we applied the Minkowski dimension code [11] to each image. This function was used to analyze the morphological complexity of the moles, after converting the images to binary and extracting their edges.

To improve the accuracy of color-based analysis in skin cancer detection, we extracted color features from various color spaces, including HSV, YCbCr, Lab, and XYZ, from the preprocessed images. This process resulted in the images ilustrated in Figures [5-8]. These features were derived using the `cv2.cvtColor` function from the OpenCV library. The analysis followed a systematic approach: (i) reading the image, (ii) converting it to the desired color spaces, (iii) applying a mask to remove dark (black) regions representing the background, and (iv) calculating key statistical metrics - mean, median, standard deviation, minimum, and maximum for each color feature. The analysis is implemented in the Python code [11].

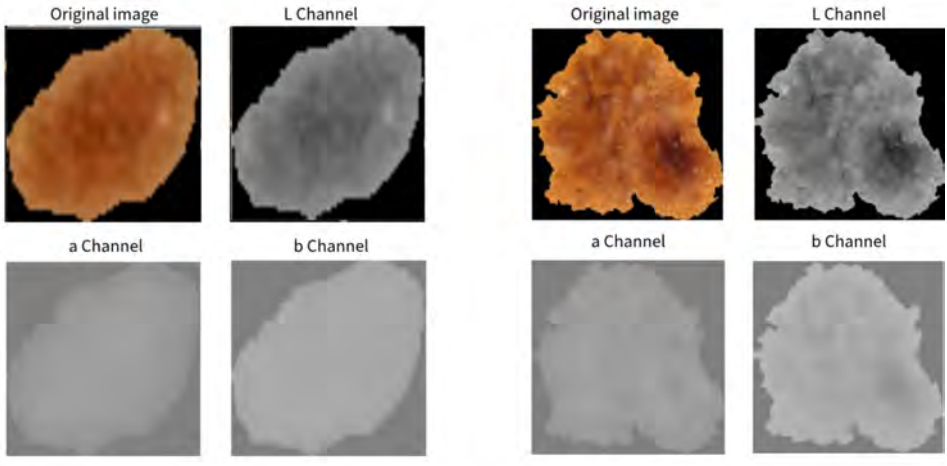


Figure 5: Lab color space layers for Benign (left) and Malignant (right) moles.

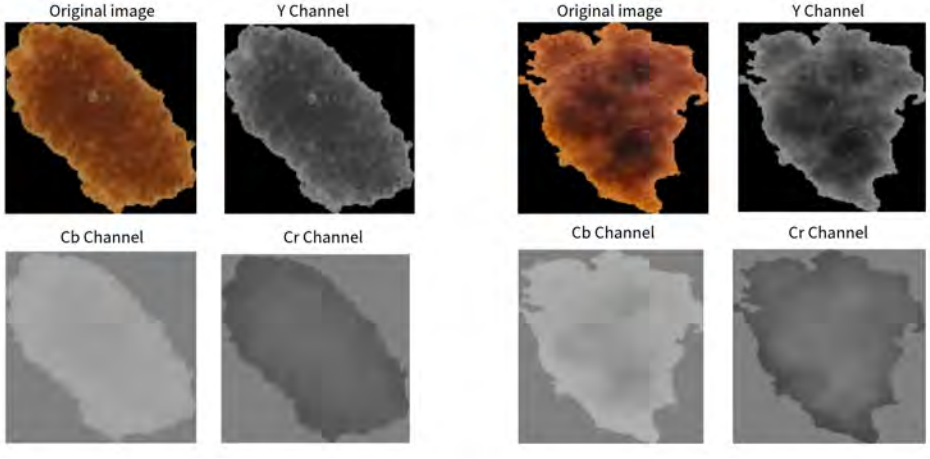


Figure 6: YCbCr color space layers for Benign (left) and Malignant (right) moles.

4 Results

Each of the images obtained via dermatoscope was processed following the procedure outlined in Section 3. The mole extraction was performed using four distinct color spaces: HSV, YCbCr, Lab, and XYZ. After extraction, the Minkowski dimension of each mole and mole's border was determined. The Minkowski dimensions of malignant moles do not follow the normal distribution when the extraction is done using LAB and YCbCr spaces. We aimed to assess whether the Minkowski dimensions of the borders of malignant and benign pigmented skin moles differ significantly. The Minkowski dimension provides a measure of the geometric complexity of the mole's border, which may vary between benign and malignant lesions. To test this, we compare the Minkowski dimensions of the two groups using statistical hypothesis testing, with Null hypothesis H_0: The Minkowski dimensions of (the borders of) malignant and benign skin moles are equal, against the

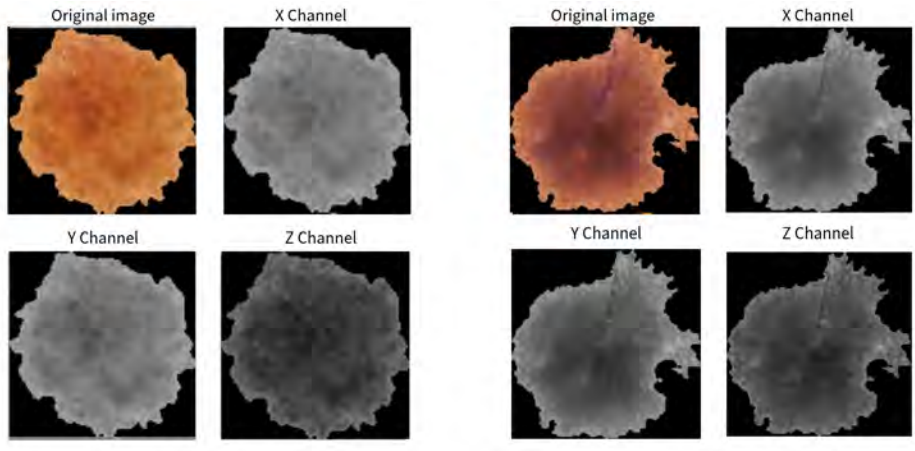


Figure 7: XYZ color space layers for Benign (left) and Malignant (right) moles.

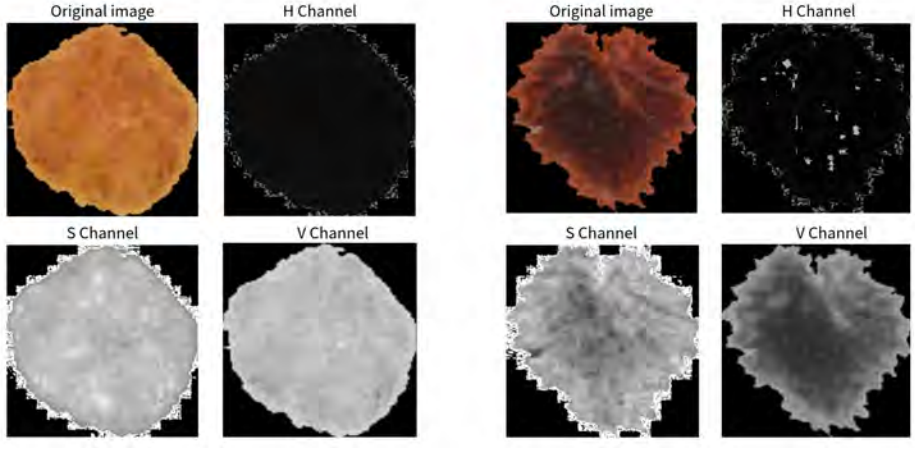


Figure 8: HSV color space layers for Benign (left) and Malignant (right) moles.

alternative hypothesis with $\alpha = 0.05$. The performed hypothesis tests gave the results presented in Table 1. Each test resulted with p-value $< \alpha$, which means that there is a significant difference between the malignant and benign moles when the Minkowski dimension is considered for both the border of the mole, and the mole itself.

Dimensions of the	Color space	HSV	YCbCr	Lab	XYZ
skin moles borders	p-value	0.000	0.000	0.000	0.000
skin moles	p-value	0.030	0.045	0.006	0.030

Table 1: Results from the hypothesis testing.

The results show that almost 76% of the malignant moles have higher Minkowski dimension of their border then the border of the benign moles. Best results are obtained if the extraction is done using YCbCr color space (75.94% of the malignant moles have

higher dimension of the border than the benign once). The weakest performance is obtained with Lab color space (73.68%). The results are significantly weaker when the whole mole is analyzed, 38.05% of the malignant moles have higher dimension than the benign moles across YCbCr space (37.05% across XYZ and HSV, and 33.75% across Lab).

Additionally, we analyzed the color of the moles in HSV, Lab, YCbCr, and XYZ color spaces. The results indicated a statistically significant difference in the minimum saturation as well as in the average saturation of the mole's pigment. Namely, benign moles tend to be more saturated than the malignant (p-value=0.000). Furthermore the standard deviation of the saturation is higher (p-value=0.021) for malignant moles as their color varies. This supports the hypothesis that color homogeneity (higher, more uniform saturation) is associated with benign moles, while color irregularity (lower mean saturation, higher variability) may be indicative of malignant moles.

When we analyze the lightness channel, we observe a statistically significant difference in the standard deviation of the lightness across the color spaces HSV, Lab, and XYZ, i.e., the deviation is higher for malignant moles. Similarly, the results suggest significant difference for minimal and maximal lightness between malignant and benign moles in the considered color spaces. When YCbCr space is used, we found that the maximal Cb component is significantly higher in malignant moles, as well as the average Cr component. Similar results are observed for the a and b channel in Lab color space.

5 Discussion and Conclusion

In this paper, we analyzed 100 dermatoscopic images of pigmented moles to examine their geometrical shape and color as part of the ABCDE diagnostic criteria. For the analysis of geometrical shape, we used the Minkowski fractal dimension, while for color evaluation, we analyzed the images across different color spaces. Specifically, we assessed saturation (HSV), lightness (HSV, XYZ, YCbCr, and Lab), and the red and blue components (YCbCr). The results showed that these features differ significantly and can therefore be used in more complex diagnostic systems.

6 Acknowledgments

This paper was partly supported by the project "NIP-sa-02-1902/8-Interdisciplinary application of data processing" financed by FEEIT Skopje, UKIM, R.N. Macedonia.

References

[1] Alfed, N; Khelifi, F. Bagged textural and color features for melanoma skin cancer detection in dermoscopic and standard images. *Expert Syst Appl.*, 90:101–10, 2017.

[2] Barata, C; Ruela, M; Mendonça, T; Marques, J.S. A bag-of-futures approach for the classification of melanomas in dermoscopy images: the role of color and texture descriptors. In:*Computer vision techniques for the diagnosis of skin cancer*. Springer, Berlin, Germany, 2014.

[3] Oukil, S; Kasmi, R; Mokrani, K; Garcia-Zapirain, B. Automatic segmentation and melanoma detection based on color and texture features in dermoscopic images. *Skin Res Technol.*, 28(2):203–211, 2022.

[4] Chen, J; Stanley, R.J; Moss, R.H; Van Stoecker, W. Colour analysis of skin lesion regions for melanoma discrimination in clinical images. *Skin Res Technol.*, 9(2):94–104, 2003

[5] Darkovska, M; Mizhimakoski, I; Ristovska Dimitrovska, A; Andova, V. Exploring Pigmented Skin Mole Borders: Minkowski Dimensions in Malignancy Detection. In *Proceedings of ETAI*, Struga, Macedonia, 2024.

[6] De Pinto, G; Mignozzi, S; La Vecchia, C; Levi, F; Negri, E; Santucci, C. Global trends in cutaneous malignant melanoma incidence and mortality. *Melanoma Res.*, 34(3):265–275, 2024.

[7] FotoFinder Dermatoscope, (https://www.fotofinder.de/en/technology/total-body-dermoscopy/bodystudio-atbm/master)

[8] Jin, Q. DermNet NZ, ABCDEs of Melanoma https://dermnetnz.org/topics/abcdes-of-melanoma, downloaded: October 4 th 2024.

[9] Majumder, S; Ullah, M. A; Feature extraction from dermoscopy images for an effective diagnosis of melanoma skin cancer. *Paper presented at: 2018 10th International Conference on Electrical and Computer Engineering (ICECE)*, Dhaka, Bangladesh, 2018.

[10] Mandelbrot,B., *The fractal geometry of nature*, W. H. Freeman and Comp., USA, 1983.

[11] MiaDark, Skin-Mole-Analysis—Border-and-Color, https://github.com/MiaDark/Skin-Mole-Analysis—Border-and-Color

[12] Moldovanu, S; Damian Michis, F.A; Biswas, K.C; Culea-Florescu, A; Moraru, L. Skin Lesion Classification Based on Surface Fractal Dimensions and Statistical Color Cluster Features Using an Ensemble of Machine Learning Techniques. *Cancers*, 13, 5256, 2021.

[13] Romero-Morelos, P; Herrera-López, E; González-Yebra, B. Development, Application and Utility of a Machine Learning Approach for Melanoma and Non-Melanoma Lesion Classification Using Counting Box Fractal Dimension. *Diagnostics*, 14, 1132, 2024.

[14] Pillay, V; Viriri, S. Skin Cancer Detection from Macroscopic Images, *In 2019 Conference on Information Communications Technology and Society (ICTAS)*, Durban, South Africa, 2019.

[15] Santos, A.R.F; Aires, K.R.T; Veras, R.M.S. Aspects of Lighting and Color in Classifying Malignant Skin Cancer with Deep Learning. *Appl. Sci.* 14, 3297, 2024.

[16] Sung, H; Ferlay, J; Siegel, R.L; Laversanne, M; Soerjomataram, I; Jemal, A; et al.. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* 71:209–249, 2021.

ECG-Derived Respiration on the Fantasia Dataset using the Signal Processing RRest Toolbox

Lana Dominković[1], **Biljana Mileva Boshkoska**[1,2], **Aleksandra Rashkovska**[1,2]
[1]Faculty of Information Studies, Ljubljanska cesta 31a, 8000 Novo Mesto, Slovenia
[2]Jožef Stefan Institute, Jamova cesta 39, 1000 Ljubljana, Slovenia
lana.caldarevic@student.fis.unm.si
{biljana.mileva, aleksandra.rashkovska}@fis.unm.si

Abstract. *In this paper, we establish baseline metrics for models that estimate respiration from electrocardiogram (ECG) signals, known as ECG-derived respiration (EDR). EDR aims to extract respiratory information from ECG data, to enable non-invasive respiratory monitoring without the need for additional sensors. To establish reliable baselines, we evaluate traditional, well-established signal processing algorithms from the RRest toolbox–a MATLAB tool designed for EDR signal processing, on the Fantasia benchmark dataset–a collection of ECG and respiration recordings obtained while watching the movie Fantasia. These baseline metrics serve as benchmarks for assessing future advancements in EDR methods. Our results highlight the effectiveness of specific algorithms, providing a foundation for the development of more accurate and robust EDR models that can enhance respiratory health monitoring.*

Keywords. ECG-derived respiration, biosignal analysis, signal processing, fantasia, respiratory signal estimation, healthcare monitoring

1 Introduction

Electrocardiogram (ECG)-derived respiration (EDR) is an emerging area of research within the broader field of biosignal analysis. EDR aims to extract respiratory information from ECG signals, which are primarily used to monitor heart activity. By leveraging ECG signals to derive respiratory data, EDR can offer valuable insights into respiratory health without requiring additional sensors, thus simplifying patient monitoring. This is particularly advantageous for continuous health monitoring, where reducing sensor complexity while maximizing the quality of gathered data is critical for both in-hospital care and everyday healthcare applications.

The drive for ECG-derived respiration arises from the increasing demand for comprehensive yet cost-effective healthcare solutions. Multi-functional body sensors, which can capture several physiological signals simultaneously, represent a major advancement in personalized healthcare [17]. However, accurately extracting meaningful information from integrated biosignals, such as deriving respiration from ECG, remains a technical challenge. Solving this issue could significantly improve clinical decision-making and

provide more accurate diagnostic tools for healthcare professionals.

In this study, we focus on setting reliable baselines using traditional signal processing methods for the problem of deriving respiratory signals from ECG. These baselines will serve as benchmarks against which we can measure the performance of future models and other approaches, ensuring that improvements are both tangible and quantifiable.

Signal processing techniques have long served as fundamental tools for analyzing and interpreting data across many domains [18]. Techniques such as Fourier Transform [10], Wavelet Transform [1], and Principal Component Analysis (PCA) [4] have provided time-tested methods for understanding and manipulating signal data. By applying these methods to our dataset, we aim to establish performance metrics that provide a comprehensive understanding of the patterns and structures inherent in the data.

The methods we employ are part of the RRest toolbox [5] – a comprehensive, publicly available toolkit designed for estimating respiratory signals and respiratory rate (RR) from ECG and photoplethysmogram (PPG) signals. The toolbox comprises over 300 algorithms that combine different signal processing techniques, enabling reproducible and systematic comparisons of RR estimation methods. By providing a standardized framework, the toolbox advances research by offering a consistent basis for evaluating the performance of different algorithms. For this work, we use the Fantasia dataset, which is publicly available via PhysioNet [7]. The baselines established using the RRest toolbox will serve as reference points for evaluating the performance of future methods in this domain.

The rest of the paper is structured as follows: in Section 2, we describe the experimental data, the investigated signal-processing algorithms, and the experimental settings. In Section 3, we present and discuss the results. We conclude with the final remarks and directions for future work in Section 4.

2 Materials and Methods

2.1 Dataset Description

We leveraged the Fantasia dataset to establish baseline performance for respiratory signal analysis using traditional signal processing techniques. The Fantasia dataset, made publicly available by PhysioNet [7], includes long-term ECG and respiration recordings from 40 healthy individuals, split equally between younger adults (ages 21-34) and elderly adults (ages 68-85). Each subject was monitored for approximately two hours in a stable supine position while watching the movie Fantasia, creating controlled conditions for EDR analysis. The ECG signals were recorded at 250 Hz, providing high temporal resolution for comprehensive analysis. Examples of simultaneously measured ECG and respiration signals are shown in Fig. 1.

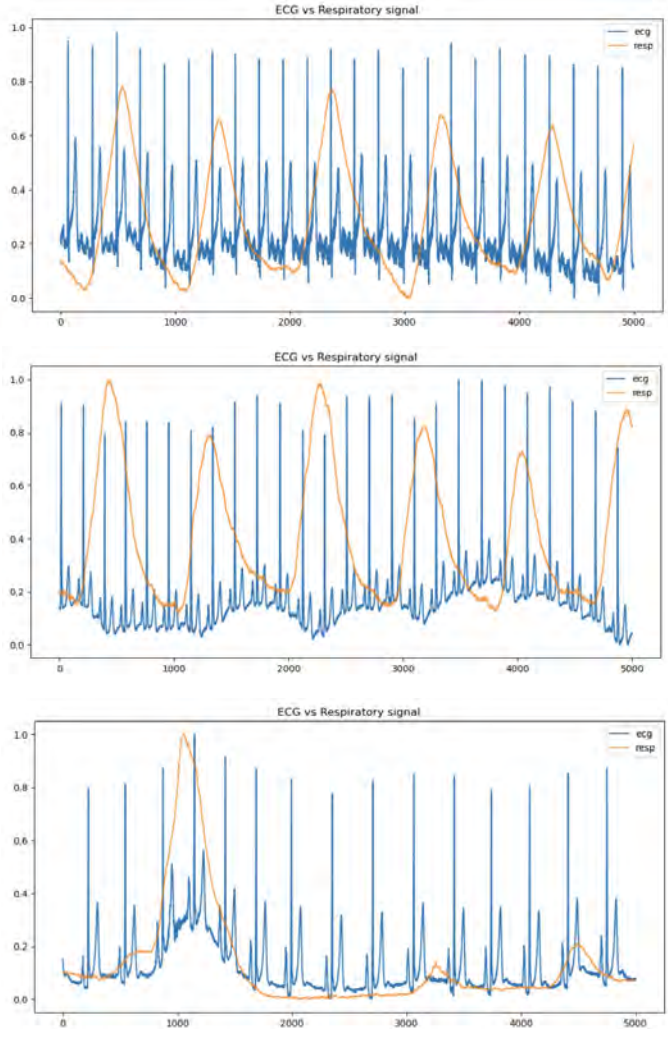


Figure 1: Examples of ECG and corresponding respiratory signal

2.2 Description of the RRest Algorithms

The RRest Toolbox [5] is a set of algorithms designed to estimate respiratory rate from physiological signals, such as ECG and PPG, implemented in MATLAB. One of the key steps in respiratory rate estimation is the extraction of respiratory signals from raw physiological data, which is crucial for deriving accurate respiratory patterns. The process begins with the removal of very low frequencies through a high-pass filter, set to a -3 dB cutoff frequency of 4 breaths per minute (bpm). Depending on the chosen extraction technique, whether filter-based or feature-based, intermediate steps vary and are explained below. The final stage, consistent across all techniques, involves applying a band-pass filter to the extracted signal, using -3 dB cutoff frequencies between 4 and 60 bpm, to eliminate non-respiratory frequencies. In continuation we explain the used algorithms.

- **ELF_RSlinB_FMeam_FPt_RDtGC_EHF**
  This algorithm extracts features related to amplitude modulation (AM) from ECG signal. It begins by eliminating low-frequency noise with a high-pass filter, then resamples the signal using linear interpolation to ensure uniform intervals. Amplitude modulation is measured by analyzing the signal's amplitude variations, with the maximum amplitude of the continuous wavelet transform (CWT) using the Morlet wavelet [2]. Peaks and troughs are identified, and their differences represent AM in the respiratory cycle [8]. Fiducial points like R-spikes and QRS complexes are detected to quantify cardiac features [14], and R-spike detection helps with timing and respiratory sinus arrhythmia extraction [16]. High-frequency noise is finally removed using a low-pass filter.

- **ELF_RSlinB_FMebw_FPt_RDtGC_EHF**
  Similar to the first algorithm. However, it extracts features based on bandwidth instead of amplitude modulation [11].

- **ELF_RSlinB_FMefm_FPt_RDtGC_EHF**
  This algorithm follows the same structure as the previous two. It extracts features based on frequency modulation (FM) rather than amplitude modulation or bandwidth [2].

- **flt_BFi**
  This filter-based algorithm applies a band-pass filter [11] to extract respiratory signals from the ECG data.

- **flt_Wam**
  A filter-based algorithm that extracts a respiratory signal corresponding to amplitude modulation using the continuous wavelet transform [2].

- **flt_Wfm**
  A filter-based algorithm that extracts a respiratory signal corresponding to fFrequency modulation using the continuous wavelet transform [2].

2.3 Performance Metrics

The performance of the algorithms in the RRest toolbox was assessed using two metrics: Mean Squared Error (MSE) and Mean Cross-Correlation (CC). MSE measures the average of the squares of the errors, which is the average squared difference between the estimated values and the actual value. Lower MSE values indicate better performance. CC measures the similarity between two signals. Higher CC values indicate that the estimated signal is more similar to the reference signal, signifying better performance.

Using the RRest toolbox, we extracted the entire respiratory signal from the complete ECG recording, employing the methods mentioned above. To calculate the Mean Squared Error (MSE) and mean Correlation Coefficient (CC), we used settings from the study conducted by [12] to process ECG and respiration signals. These parameters ensure that our calculated metrics are consistent and comparable for future research. Specifically, we employed 32-second length windows with 50% overlap, normalized the signals to a [0, 1] range, and downsampled it to 5Hz, as described in [12].

3 Results and Discussion

The performance of the RRest signal processing methods is shown in Table 1. The results show that the feature-based algorithms perform better than the filter-based, which is in agreement with the findings made in the original RRest study [5] and the study made in [12]. From the feature-based algorithms, the AM-based method named: "ELF_RSlinB_FMeam_FPt_RDtGC_EHF", achived the highest mean CC (0.59) and a relatively low mean MSE (0.073). This result confirms that EDR algorithms based on the amplitude variations of the ECG waves are promising direction, which is consistent also with previous findings in EDR research [18]. However, there is still room for improvement, as the overall cross-correlation values suggest that the signal reconstruction is not fully optimized.

Method	**Mean CC**	**Mean MSE**
ELF_RSlinB_FMeam_FPt_RDtGC_EHF	0.59	0.073
ELF_RSlinB_FMebw_FPt_RDtGC_EHF	0.50	0.069
ELF_RSlinB_FMefm_FPt_RDtGC_EHF	0.56	0.070
flt_BFi	0.37	0.083
flt_Wam	0.38	0.093
flt_Wfm	0.44	0.081

Table 1: Performance of RRest signal processing methods on the Fantasia dataset

4 Conclusions

This study provides baseline performance metrics for ECG-derived respiration (EDR) using traditional signal processing methods. The findings show that, while the AM-based method achieved a relatively high cross-correlation, the current performance remains insufficient for reliable use in clinical practice. The cross-correlation values observed are not yet satisfactory to support accurate, simultaneous extraction of ECG and respiratory information from a single biomedical signal. This indicates room for improvement, as further refinement of EDR techniques is essential to meet the standards required for medical applications. Addressing these technical challenges is critical for advancing cost-effective, multi-functional biosensors that can enhance continuous health monitoring and aid clinical decision-making.

In future work, we plan to explore the use of deep learning architectures such as modified UNet, which has shown promising results in previous studies for extracting respiratory signals from ECG data. Deep learning models are well-suited for capturing complex, non-linear patterns within time-series data and hold potential for improving accuracy, minimizing MSE, and enhancing the similarity between the derived and reference signals. By integrating these advanced techniques, we aim to develop more reliable methods for non-invasive respiratory monitoring, further advancing the field of ECG-derived respiration analysis. In addition to leveraging the Fantasia dataset, exploring other publicly available datasets for ECG-derived respiration could provide valuable insights and improve model generalization.

5 Acknowledgments

Author Aleksandra Rashkovska acknowledges the financial support from the Slovenian Research and Innovation Agency (ARIS) under Grant No. P2-0095. Biljana Mileva Boshkoska acknowledges the financial support from ARIS under Grant No. P2-383.

References

[1] Addison, P. S. Wavelet transforms and the ECG: A review. *Physiological Measurement*, 26(5):R155-R199, 2005.

[2] Addison, P. S; Watson, J. Secondary transform decoupling of shifted nonstationary signal modulation components: Application to photoplethysmography. *International Journal of Wavelets, Multiresolution and Information Processing*, 2(1):43-57, 2004.

[3] Aqajari, S. A. H; Cao, R; Zargari, A. H. A; Rahmani, A. M. An end-to-end and accurate PPG-based respiratory rate estimation approach using cycle generative adversarial networks. In *Proceedings of the 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 2029-2032, 2021.

[4] Castells, F; Laguna, P; Sornmo, L. Principal component analysis in ECG signal processing. *EURASIP Journal on Advances in Signal Processing*, Article number: 074580, 2007.

[5] Charlton, P. H; Bonnici, T; Tarassenko, L; Clifton, D. A; Beale, R; Watkinson, P. J. An assessment of algorithms to estimate respiratory rate from the electrocardiogram and photoplethysmogram. *Physiological Measurement*, 37(4):610, 2016.

[6] Gauthier, D. J; Bollt, E. M; Griffith, A; Barbosa, W. A. S. Next generation reservoir computing. *Nature Communications*, 12(1):1-8, 2021.

[7] Goldberger, A. L; Amaral, L. A; Glass, L; Hausdorff, J. M; Ivanov, P. C; Mark, R. G; Mietus, J. E; Moody, G. B; Peng, C. K; Stanley, H. E. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215-e220, 2000.

[8] Karlen, W; Raman, S; Ansermino, J. M; Dumont, G. A. Multiparameter respiratory rate estimation from the photoplethysmogram. *IEEE Transactions on Biomedical Engineering*, 60(7):1946-1953, 2013.

[9] Karlen, W; Turner, M; Cooke, E; Dumont, G; Ansermino, J. M. CapnoBase: Signal database and tools to collect, share and annotate respiratory signals. In *Proceedings of the Annual Meeting of the Society for Technology in Anesthesia (STA 2010)*, pages 25-25, West Palm Beach, FL, USA, 2010.

[10] Kumar, A; Ranganatham, R; Komaragiri, R; Kumar, M. Efficient QRS complex detection algorithm based on Fast Fourier Transform. *Biomedical Engineering Letters*, 9(1):145-151, 2018.

[11] Lindberg, L. G; Ugnell, H; Oberg, P. A. Monitoring of respiratory and heart rates using a fibre-optic sensor. *Medical and Biological Engineering and Computing*, 30(5):533-537, 1992.

[12] Merdjanovska, E; Rashkovska, A. Respiration extraction from single-channel ECG using signal-processing methods and deep learning. In *Proceedings of the 2020 43rd International Convention on Information, Communication and Electronic Technology (MIPRO)*, pages 848-852, Opatija, Croatia, 2020.

[13] Pimentel, M. A. F; et al. Towards a robust estimation of respiratory rate from pulse oximeters. *IEEE Transactions on Biomedical Engineering*, 64(8):1914-1923, 2016.

[14] Rajkumar, K; Ramya, K. Respiration rate diagnosis using single lead ECG in real time. *Global Journal of Medical Research*, 13(7):7-11, 2013.

[15] Ravichandran, V; et al. RespNet: A deep learning model for extraction of respiration from photoplethysmogram. *arXiv preprint*, arXiv:1902.04236, 2019.

[16] Ruangsuwana, R; Velikic, G; Bocko, M. Methods to extract respiration information from ECG signals. In *ICASSP 2010 - IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 570-573, Dallas, Texas, USA, 2010.

[17] Trobec, R; Avbelj, V; Rashkovska, A. Multi-functionality of wireless body sensors. *The IPSI BgD Transactions on Internet Research*, 10:23-27, 2014.

[18] Trobec, R; Rashkovska, A; Avbelj, V. Two Proximal Skin Electrodes - A Respiration Rate Body Sensor. *Sensors*, 12:13813-13828, 2012.

Factors of the development and manifestation of the talent of managers

Vesna Stepišnik, Janez Kolar
School of Advanced Social Studies,
Gregorčičeva 19, 5000 Nova Gorica, Slovenia
Vesna.Stepisnik@gmail.com; Janez.Kolar@fuds.si

Abstract: *In this article, we present our research, in which we will investigate the influence of 25 environmental, intrapersonal and interpersonal factors on the development of the talent of managers, who were recognized as key in the development and manifestation of the talent of managers in the context of the literature review. The research will be based on already realized talents. We will follow the guidelines of some scientists [3], [8], who call for research (instead of "recipes") how managers themselves perceive and experience their role and how they determine "what makes a good manager and what he actually wants to be" [18, 417-438]. Based on the TAD model (Talent development in achievement domain) [17], we will also investigate how factors affect the levels (ability, competence, expertise, eminence) of talent development of managers.*

Key Words*: talent management, development factors, model, levels of talent development*

1 Introduction

New, rapidly emerging and changing forms of work increasingly express the need for many skills (such as, for example, complex thinking, quick response, the ability to think causally, empathy, emotional and social intelligence), which are often associated with talents in individual fields. There will be a growing need for talents who will be responsible for the development of their own careers [13]. In addition to education and experience, the modern needs of companies are increasingly oriented towards the use of soft skills, which can be associated with talent [19], [1]. For the transition to better »soft approaches« to the discovery, placement and development of talents in society, a transition to the formation of education and employment policies will be necessary, which will enable the change in the paradigmatic approach of society from »satisfaction with employment as such« to » self realization« in education and employment.

Throughout history, researchers have defined talent differently. Some researchers who studied talent tended to define that talent is genetically determined, others that it is learned. The initial literature on giftedness [7], [20] defined as gifted children those who achieved high IQ results, later some researcher [5], [6] opposed the issue. Nijs et al. [14, 13] formulated a definition of talent in their research: »Talent refers to the systematically developed innate abilities of individuals who are deployed in activities they like, find important and in which they want to invest energy«.

Through the history of the scientific study of talents, two main approaches have emerged. The exclusive approach, which is still dominant, is aimed at the support and development of individuals, but some scientists [21] encourage an inclusive approach that allows and encourages the discovery and development of everyone's talent. In the last few years, a new discourse has emerged in the field of studying talents - a multidimensional approach

that emphasizes the importance of context and development when defining talents. »On the basis of a transactional perspective, giftedness is based on a transactional entity of an individual profile« [16, 172-184].
Jarvin and Subotnik [12], Preckel et al. [17] note that in the process of talent development, talent represents the initial phase, when an individual acquires knowledge, experience and skills through learning and practice, then talent grows into the competence phase, then into the expertise phase and then into the eminence phase. Also, Heller et al. [11] after Bezić [2, 13] note, that »during the development of talent, the process is influenced by internal (personal) and external (environmental) factors, which they call catalysts«. Some people can succeed at a certain level, but when they advance to the next level, their effectiveness declines. Some researchers [11], [22], note, that progress is possible in the field of innate abilities, dexterity and skills in an individual, which reflects the development (talent adds knowledge and experience) significantly when achieving and overcoming tasks. Such a person can add innovation, his own creative note to some field, knowledge and experience. Creativity appears as a resource of identification. We are talking about excellence.
Slovenia has a poorly developed management of talented personnel. The international preliminary study of personnel practices CRANET 2015 showed that Slovenia lags behind other European countries in terms of development, diversity and intensity of use of personnel practices. Research shows that companies are more focused on managing talent than on working with talent (learning practices in use of tools that enable the development of talent itself). While the school system includes the recognition and development of talents, for the adult population in Slovenia (except for some large companies) we do not have customized approaches and systems for the discovery and development of talents. Zobarič, Kastelec and Podnar [22] note, that the task of a successful manager is primarily work and relationship with employees. On the other hand, a successful manager has a wide range of competencies. Because some of them are instructive, some are not, unfortunately, not all managers can be successful. Some researcher [15], [9] find an association between the impact of managers personality traits and personal efficiency and team performance.

2 Methodology

In order to gain insight into the situation from different perspectives, we decided to conduct the research using a mixed method Fig 1.
As part of a quantitative survey of 100 to 200 managers of the most successful medium sized organizations in Slovenia, we will use a cross-sectional survey that includes data collection over a period of time. We will investigate which of the 25 preliminary talent development factors are related to the level (abilities, competences, expertise, eminence) of the manager's talent development using the TAD model [17]. Fig. 2.
The results obtained on the basis of quantitative analysis will be checked with a qualitative analysis of 3 - 5 managers, each of whom visibly results in his own success in his work. We will also investigate which level of talent development based on the TAD model [17] the managers fall into.

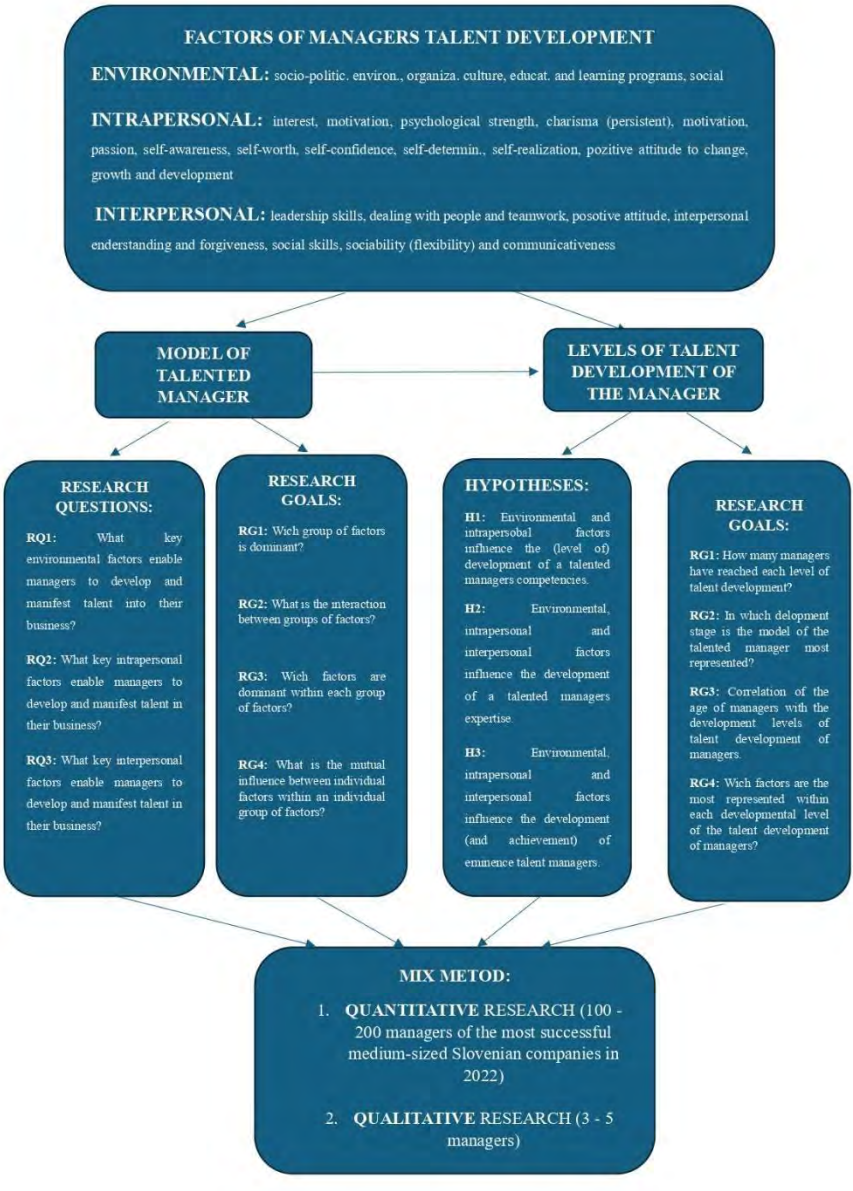


Figure 1: Research plane

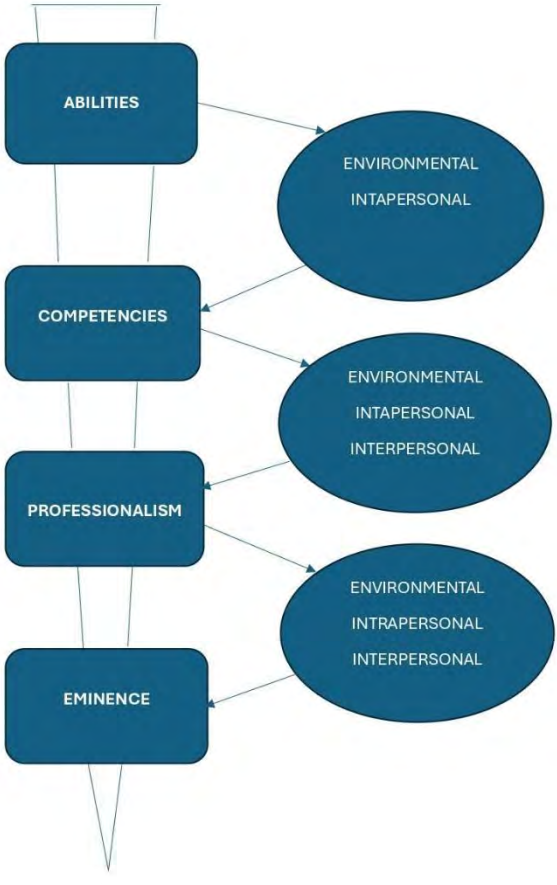


Figure 2: The influence of factors on the levels of talent development

3 Conclusion

"The World Economic Forum recent study on "Future of Jobs 2020" shows that 90.3% of companies will provide more remote working opportunities and 87.1% are accelerating the working process's digitalization. 58.1% of organizations are accelerating tasks' automation, 67% are more likely to hire new temporary staff with skills relevant to new technologies, and 65% of organizations are likely to outsource some business functions to external contractors" [4, 280-289].

In the labor market, managers play an important role in the strategic decisions of companies and the development of the economy, therefore the selection of future managers is one of the key steps in the development of society, especially at a time when we are increasingly short of staff on the labor market.

With the help of IT, it will be necessary to create platform with different applications based on the original model of talent management development factors, which we will design based on the results of the research, as a tool for introducing the model into practice.

This applications based on original model of a talented manager will be able to:

- helped companies in the search, selection and development of talent managers, in the conceptualization of management development and succession planning;
- served as a tool (aid) in the creation of training programs (with an emphasis on skills) for future managers;
- served in the design of programs to encourage the development of young talents, in further research and in the design of education policies, which will be aimed at the entire children's perspective;
- enable HR departments to include the model as a self-assessment tool in their talent identification and status processes;
- helped to shed light on components of talent that are not always visible to other parties. As a tool (curriculum), the model will also help to correlate the candidate's self-management skills and his education.
- By creating a model of talent manager excellence, we will help create a clearer (still inconsistent) definition of talent.

4 References

[1] Babaeva, Julia D, Alexander E. Voiskounsky. 2002. IT-Giftedness in Children and Adolescents. Journal of Educational Technology & Society 5 (1): 154–62

[2] Bezić, Tanja. 2019. Professional starting points for the modernization of the concept of discovering gifted slaves, scholars and students and educational and educational work with them. Ljubljana: Institute of the Republic of Slovenia for Education.

[3] Bresnen, Mike, Damian Hodgson, Simon Bailey, John Hassard in Paula Hyde. 2019. Hybrid Managers, Career Narratives and Identity Work: A Contextual Analysis of UK Healthcare Organizations. Human Relations 72 (8): 1341–68.

[4] Dutta, Debolina. Prem Mirchandani, K. P. Anasha Dutta, D., Mirchandani, P., & Anasha, K. P. (2021). Internal GIG Ecosystem in IT/ITeS Organizations for Nonlinear Growth. South Asian Journal of Human Resources Management, 8(2), 280-289.

[5] Ericsson, K. Anders, Ralf T. Krampe in Clemens Tesch-Röme. 1993. The Role of Deliberate Practice in the Acquisition of Expert Performance. Psychological Review 100 (3): 363–406.

[6] Gallardo-Gallardo, Eva, Nicky Dries in Tomás F. González-Cruz. 2013. What Is the Meaning of 'Talent' in the World of Work? Human Resource Development Review 23 (4): 290–300.

[7] Garrison, C. G., A. Burk in L. S. Hollingworth. 1917. Psychology of a Prodigious Child. Journal of Applied Psychology, 1 (2): 101–10.

[8] Gjerde, Susann in Mats Alvesson. 2020. Sandwiched: Exploring Role and Identity of Middle Managers in the Genuine Middle. Human Relations 73 (1): 124–51.

[9] Griffin, Mark A., Andrew Neal in Sharon K. Parker. 2017. A New Model of Work Role Performance: Positive Behavior in Uncertain and Interdependent Contexts. The Academy of Management Journal 50 (2): 327–47.

[10] Harding, Nancy, Hugh Lee in Jackie Ford. 2014. Who Is ‚the Middle Manager'? Human Relations 67 (10): 1213–37.

[11] Heller, Kurt A., Franz J. Mönks, Robert J. Sternberg in Rena F. Subotnik, ur. 2000. *International Handbook of Giftedness and Talent*. Amsterdam: Elsevier.

[12] Jarvin, Linda in Rena F. Subotnik. 2010. Wisdom From Conservatory Faculty: Insights on Success in Classical Music Performance. Roeper Review 32 (2): 78–87.

[13] Kohont, Andrej. 2019. Talent management. Ljubljana: Faculty of Social Sciences, FDV Foundation.

[14]Nijs,Sanne, Eva Gallardo-Gallardo, Nicky Dries in Luc Sels. 2014. A Multidisciplinary Review into the Definition, Operationalization, and Measurement of Talent. Journal of World Business 49 (2): 180–191.

[15] Novak, Igor. 2021. The leader perceives the connection between personal characteristics and personal performance in team performance. Theory and practice, 1048–64

[16] Owen, Lo C., Marion Porath, Yu Hsiao-Ping, Chen Chen-Ming in Wu I-Chen. 2018. Giftedness in the Making: A Transactional Perspective. Gifted Child Quarterly, 172– 184.

[17] Preckel, Franzis, Jessika Golle, Roland Grabner, Linda Jarvin, Aaron Kozbelt, Daniel Müllensiefen, Paula Olszewski-Kubilius idr. 2020. Talent Development in Achievement Domains: A Psychological Framework for Within- and Cross-Domain Research. Perspectives on Psychological Science 15 (3): 691–722

[18] Rostron, Ali. 2021. How to Be a Hero: How Managers Determine What Makes a Good Manager through Narrative Identity Work. Management Learning 53 (3): 417– 438.

[19] Silverman, Linda K. 1982. The gifted and talented. V *Exceptional children and youth: An introduction*, ur. Edvard L. Meyen, 168–95. London: Love Publishing Company.

[20] Warne, Russell T. 2016. Five Reasons to Put the g Back Into Giftedness: An Argument for Applying the Cattell-Horn-Carroll Theory of Intelligence to Gifted Education Research and Practice. Gifted Child Quarterly 60 (1): 3–15.

[21] Ziegler, Albert, Heidrun Stoeger in Wilma Vialle. 2012. Giftedness and Gifted Education: The Need for a Paradigm Change. Gifted Child Quarterly 56 (4): 194–197.

[22] Zobarič, Martina, Patricija Kastelec in Nataša Podnar. 2020. Competence of well-being managers. Review for universal excellence 9 (1).

Application of prediction models in educational context

Laura Fink
GEA College, Faculty of Entrepreneurship
WTC, Dunajska cesta 156, 1000 Ljubljana
{laura.fink@gea-college.si}

Abstract:
The debate in Slovenia over substituting the results of students who miss the external exam with a statistically calculated estimate based on peers' average results with similar academic achievement has led to this review of prediction models of student achievement. The discussion is based on the decree on the criteria for the selection of candidates into the secondary school programs and the associated rules on the external national examination, which were adopted in August 2024. The purpose of this study is to present an overview of research on prediction models of student achievement in the educational context, explore their performance, identify frequent applications, and provide examples of their use. Additionally, our research aims to contribute to a scientific discourse about the questionable meaningfulness and eligibility of using estimates obtained from the prediction models for result substitution. The use of prediction models is primarily aimed at implementing timely measures, enabling teachers to intervene promptly, prevent student dropouts, and reduce failure rates. On the other hand, to the best of our knowledge, the use of prediction results to substitute individual outcomes in case of their absence from external examination is not yet explored and therefore requires thoughtful consideration. We elaborate on the concerns that such practice could raise. We suggest that the results of prediction models must definitely be used with care, ensuring they prevent potentially unfavorable situations or enhance learning outcomes rather than restricting, predetermining, or absolving anyone from taking the examination. It is also questionable to substitute individual achievements with predictions based on results from a group of students with similar characteristics, as the prediction models we reviewed primarily rely on the individual students' previous achievements. Making predictions for any kind of individual predetermination is a questionable practice that requires a broader debate. This practice raises ethical, pedagogical, and motivational questions that need to be further discussed in professional and scientific discourse.

Key Words: *educational data mining, prediction accuracy, learning analytics, student academic achievement, student academic performance, model performance*

1 Introduction

The recent public debate in Slovenia [32] [19] [37] over substituting the results of students who do not attend the external exam with a statistically calculated estimate based on the average results of peers with similar academic achievements prompted this investigation into models that predict student achievement.

The decree on the "criteria for the selection of candidates in the event of restriction of

enrollment in programs of secondary vocational education, secondary professional education, and gymnasiums" [23] [25] [32] says that starting with the school year 2024/2025, the results of external national exams in Slovene language and mathematics will be used as part of the selection process. Together, they will make up to 40% of the selection criteria, or 20% each. The remaining 60% will be determined on the basis of final grades from compulsory elementary school subjects in grades 7th, 8th, and 9th.

The primary school national examination rules [24] stipulate that students who cannot attend the external national examination due to justifiable reasons will have their achievement replaced. As recapitulated by the [32], “The substitute achievement will be determined for the purposes of the enrolment procedure as the median achievement at the national examination of those candidates who have the same sum of the final grades of this subject in grades 7th, 8th, and 9th." Apart from medical reasons necessitating a medical certificate, national rules allow excusable reasons for non-participation in the national examination, such as a natural or other accident, a death in the family, or participation in an international knowledge, skills, or sports competition.

The purpose of this study is to present an overview of research on prediction models of student achievement in the educational context, explore their performance, identify frequent applications, and provide examples of their use. Additionally, our research aims to contribute to a scientific discourse about the questionable meaningfulness and eligibility of using estimates obtained from the prediction models for result substitution.

First, to predict student achievement, researchers have used a wide range of methods. Some of these include data mining [22] [26] [17] [1], neural networks [8] [27], decision trees [11] [29] [18], Bayesian-based techniques [20] [36], support vector machines [28] [6], correlation and regression, and others [38].

Second, the prediction models may include different types of the learning outcomes [26]. For example, we can measure learning outcomes using a pass/fail system, points earned, or rankings. Some models incorporate formative assessments as learning outcomes, allowing for early intervention, adjustment of teaching approaches, and feedback throughout the course. Conversely, other models incorporate summative assessments at the course's conclusion. Some models even include student self-assessment and opinions, as well as outcomes based on subjective measurements [39] like satisfaction, self-esteem, and meaning [26].

Thirdly, extensive research has been done on the input factors that predict a student's performance [7] [15] [12] [14]. The factors that are frequently used in predicting academic achievement include different demographic, academic, behavioral [3], psychological, students’ environment, students’ activity [2], and other factors. Demographic factors can include gender, nationality, and others. Academic factors can include school year, semester, course topic, student group, student grade level, and student educational stage [7]. Behavior [21] [7] can include, for example, the use of an online classroom and other learning materials, raising hands, viewing announcements, active discussion, attendance, and absence (Gottfried, 2009). Other factors include self-beliefs [41], teacher-student, parent-student, and peer relationships [46] [45], learning style [4], achievement goals [12], prior achievement [12], preparedness for a course [16], class characteristics [13], and teachers' approaches [33] [35] [31]. The list of the

aforementioned factors is certainly not entirely complete. Each researcher uses somewhat different categories of factors.

Finally, the prediction models include various contexts, ranging from classroom teaching to learning analytics in online teaching. Interestingly, Namoun and Alshanqiti [26], in their review, show that the large majority of predictions relate to computer science, mathematics, engineering, and psychology study programs.

2 Methods

This research is primarily based on a literature review of prediction models of academic achievement in the educational context. It strives to present the prerequisites of an effective prediction model and highlights their main applications.

3 Discussion

While research extensively explores factors and optimal prediction techniques and models, it is rare to find a research article that outlines the actual effects and consequences of substituting the results of absent students from external examinations with statistical estimates. Regardless of the technique used, the research papers we reviewed did not suggest or test the use of prediction to substitute student results in case of absence from the final external examination.

Rather, research (e.g., [44]) suggests the application in identifying students at-risk of failing external examinations, talented students, and potential dropouts with the aim of motivating and assisting them to develop further and pursue close to the limit of their capabilities. Gerritsen [8], for example, sees the models and algorithms that predict student performance as the "preventive instrument to decrease academic dropout." Prediction models are often utilized to streamline the decision-making process [14]. This includes selecting candidates for educational programs and enhancing learning and teaching methods [15]. These models play a crucial role in formulating intervention strategies to avert students' failure in external examinations. The schools are under pressure to decrease the failure rate and improve their students' learning outcomes, academic achievements, and future career prospects. In an effort to improve the results of their students at the external final examination, many schools use predictions and with that indirectly acknowledge that part of the responsibility for the students' academic achievement lies in pedagogical approaches, teachers' efficacy, and different learning methods, including peer learning, for example.

With an aim to improve the results of the students, schools are implementing measures to improve student outcomes, such as increasing teacher training to apply personally tailored pedagogical approaches. These approaches, which are mainly based on Vygotsky's zone of proximal development [5] [34], such as scaffolding, encourage students to progressively take responsibility for their learning outcomes and their learning path and allow them to gradually, at their own pace, develop their independence. According to

Vygotsky [42], what a student "is able to do in collaboration today he will be able to do independently tomorrow." In the literature review, we found various applications of prediction model results, but not a single study described or suggested using prediction results to substitute real data for students who missed the final external examination.

According to the research, some other good practices regarding the application of prediction results include facilitation of advice regarding placements [15] in different study programs, additional coaching, or other support offered to students. The predictions can also be useful in the early recognition of potential issues, identification of areas for improvement, and further developing learning objectives and learning outcomes [20]. Research suggests that based on the prediction models, teachers might consider adjusting the content and teaching methods, approaching the students at risks, taking additional effort in motivating them, and addressing students either individually or addressing the entire class, if relevant.

Any statistical model contains, one way or another, some limitations. When using models to predict outcomes, it is absolutely critical to acknowledge each model's limitations. Some researchers, for example, Gerritsen [8], reports replacing and omitting missing data from their dataset before performing analysis. Wongvorachan et al. [43] investigate class imbalances such as oversampling and under sampling.

The central theme of most research papers dealing with prediction models is the "evaluation of the model performance" [44], measured by "accuracy, precision, recall, F-score, and area under the ROC curve." These measures that are derived from the confusion matrix [30] [40] reflect the correctness of predictions by providing "the number of correctly and incorrectly classified instances" [44].

Table 1: The confusion matrix

		Predicted	
		Positive/True	**Negative/False**
Actual	**Positive**	True positive	False negative
	Negative	False positive	True negative

Source: [30] [40]

Different authors (e.g., [9]) tested numerous models by comparing measures such as accuracy, precision, recall, and F-score based on different methodologies and different explanatory input variables. Francis and Babu [7, p.14], for example, showed that the models based on the clustering technique "performed much better than a decision tree or a neural network," which are otherwise the most commonly applied to predict student performance [15]. Furthermore, the accuracy of their [7] models improved the most when they included a set of academic, behavioral, and other explanatory input variables (collaboration of parents, parents' satisfaction, and student absence days). While it is crucial for the model's performance what and how many explanatory input variables it includes, the accuracy of the model does not always improve by adding additional explanatory input variables. Some models with three explanatory input variables [44] perform better in terms of accuracy than models that include all the explanatory input

variables available. Some variables might contribute more to the accuracy of the model than others. Additionally, how much the accuracy improves by adding a certain variable to the model also depends on the other variables already in the model.

Models that researchers use to make predictions must meet stricter criteria compared to other analyses. The validity of using the results of models that apply to the group of students rather than individuals to substitute individual results with predictions in case of their absence raises additional concerns. Furthermore, it is important to raise the numerous considerations of Kizilcec and Lee [16] and others concerning algorithmic fairness.

4 Conclusion

The results of prediction models are commonly used to timely identify potential dropouts, students at risk of failing external examinations, and outstanding students. Research suggests that prediction models play a crucial role in determining intervention measures to prevent students' failure at the external examinations promptly. Schools are under pressure to decrease the failure rate, improve the learning outcomes and academic achievements, and ensure the future career prospects of their students. The use of prediction models is primarily aimed at implementing timely measures, enabling teachers to intervene promptly, prevent student dropouts, and reduce failure rates.

On the other hand, to the best of our knowledge, the use of prediction results to substitute individual outcomes in case of their absence from external examination is not yet explored and therefore requires careful consideration. The debate about the effects of substituting the achievements of students who cannot attend the external national examination became relevant in the light of the decree on the criteria for the selection of candidates in the event of restriction of enrollment in secondary school programs [23] and the associated rules on the external national examination [24] that were adopted in Slovenia in August 2024.

Generally, the choice of prediction model should be carefully selected based on model performance measures such as accuracy, precision, recall, F-score, and others. These performance measures are subject to change depending on the explanatory input variables included in the model and the dataset itself. However, previous research demonstrates that adding more exploratory input variables does not simply improve the model's performance. Therefore, in building the model, one should consider how much the particular input variable contributes to the power of the model; not only as a standalone variable but in different combinations with other variables. Further, one should critically reflect on the models' limitations, including variables that maybe should be but are not included in the model. Importantly, one should question the validity of substituting individual achievements with predictions that are based on the results of a group of students with similar characteristics. This is yet another questionable and risky practice, since the prediction models we reviewed are primarily based on individual students' own previous achievements aimed at promptly intervening and not substituting results. The purpose for which the results of predictive models are used is crucial.

The use of prediction models for substituting students' achievement is questionable since the estimates of predictive models are never absolutely certain. The prediction models are all about the probability that the real results would match a prediction. The performance

and the power of the prediction model improve when the probability that the real results match the prediction improves. Regardless of the technique used, the research papers we reviewed did not suggest or test the use of prediction to substitute student results in case of absence from the final external examination. The literature suggests and examines the use of prediction models to act preventively and enable them to implement timely measures, intervene promptly, prevent student dropouts, and reduce failure rates.

Based on the findings of this study, we suggest broadening thoughtful discussion, involving school principals who can provide insights into the real-world situations in schools, and exploring alternative options for the feasibility of external examinations. The results of prediction models must definitely be used with care in a way that helps prevent potentially unfavorable situations or backlogs or improves learning outcomes and not in a way that limits or predetermines anyone. Making predictions for any kind of individual predetermination is a questionable practice that requires a broader debate. The outcomes and consequences of this practice can serve as a basis for broader discussions among scientists and a deeper exploration of these questions within a wider scientific community. In the context of professional and scientific discourse, we should further address the ethical, pedagogical, and motivational questions that such practice raises.

9 References

[1] Albreiki, B., Zaki, N., & Alashwal, H. (2021). A systematic literature review of student'performance prediction using machine learning techniques. *Education Sciences*, *11*(9), 552.

[2] Alyahyan, E., & Düştegör, D. (2020). Predicting academic success in higher education: literature review and best practices. *International Journal of Educational Technology in Higher Education*, *17*(1), 3.

[3] Babić, I. D. (2017). Machine learning methods in predicting the student academic motivation. Croatian Operational Research Review, 8(2), 443–461. https://doi.org/10.17535/crorr.2017.0028

[4] Binh, H. T., & Duy, B. T. (2017). Predicting students' performance based on learning style by using artificial neural networks. *9th international conference on knowledge and systems engineering (KSE)* (p. 48-53). IEEE. https://ieeexplore.ieee.org/abstract/document/8119433

[5] Chaiklin, S. (2003). The zone of proximal development in Vygotsky's analysis of learning and instruction. In Kozulin, A. (ed.), Gindis, B. (ed.), Ageyev, V. S. (ed.), Miller, S.M. (ed.). *Vygotsky's educational theory in cultural context*, 39-64.

[6] Damuluri, S., Islam, K., Ahmadi, P., & Qureshi, N. S. (2020). Analyzing navigational data and predicting student grades using support vector machine. *Emerging Science Journal*, *4*(4), 243-252.

[7] Francis, B. K., & Babu, S. S. (2019). Predicting academic performance of students using a hybrid data mining approach. *Journal of medical systems*, *43*(6), 162.

[8] Gerritsen, L. (2017). Predicting student performance with Neural Networks. *Master Thesis*. Tilburg University, Netherlands.

[9] Ghorbani, R., & Ghousi, R. (2020). Comparing different resampling methods in predicting students' performance using machine learning techniques. *IEEE access*, *8*, 67899-67911.

[10] Gottfried, M. A. (2009). Excused versus unexcused: How student absences in elementary school affect academic achievement. *Educational Evaluation and Policy Analysis*, *31*(4), 392-415. https://journals.sagepub.com/doi/abs/10.3102/0162373709342467

[11] Hamoud, A., Hashim, A. S., & Awadh, W. A. (2018). Predicting student performance in higher education institutions using decision tree analysis. *International Journal of Interactive Multimedia and Artificial Intelligence*, *5*, 26-31.

[12] Harackiewicz, J. M., Barron, K. E., Tauer, J. M., & Elliot, A. J. (2002). Predicting success in college: A longitudinal study of achievement goals and ability measures as predictors of interest and performance from freshman year through graduation. *Journal of educational psychology*, *94*(3), 562.

[13] Helal, S., Li, J., Liu, L., Ebrahimie, E., Dawson, S., Murray, D. J., & Long, Q. (2018). Predicting academic performance by considering student heterogeneity. *Knowledge-Based Systems*, *161*, 134-146.

[14] Huang, S., & Fang, N. (2013). Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models. *Computers & Education*, *61*, 133-145. https://www.sciencedirect.com/science/article/abs/pii/S0360131512002102

[15] Jantakun, K., Jantakun, T., & Jantakoon, T. (2022). The architecture of system for predicting student performance based on data science approaches (SPPS-DSA architecture). *International Journal of Information and Education Technology*, *12*(8),

778-785.
[16] Kizilcec, R. F., & Lee, H. (2022). Algorithmic fairness in education. In Holmes, W., Porayska-Pomsta, K. (Eds.), *The ethics of artificial intelligence in education*, Routledge. p. 174-202.
[17] Khan, A., & Ghosh, S. K. (2021). Student performance analysis and prediction in classroom learning: A review of educational data mining studies. *Education and information technologies*, 26(1), 205-240.
[18] Kolo, D. K., & Adepoju, S. A. (2015). A decision tree approach for predicting students academic performance. *International Journal of Education and Management engineering, 5*, 12-19
[19] Kuralt, Š. (2024). Pri omejitvi vpisa bodo ocene štele največ 60odstotkov. Delo. https://www.delo.si/novice/slovenija/pri-omejitvi-vpisa-bodo-ocene-stele-najvec-60-odstotkov
[20] Makhtar, M., Nawang, H., and Shamsuddin, S. N. W. (2017). Analysis on students' performance using Naïve Bayes classifier. *Journal of Theoretical and Applied Information Technology 95*(16), 3993–3999.
[21] Mavroveli, S., & Sánchez-Ruiz, M. J. (2011). Trait emotional intelligence influences on academic achievement and school behaviour. *British Journal of Educational Psychology*, *81*(1), 112-134.
[22] Minaei-Bidgoli, B., Kashy, D. A., Kortemeyer, G., & Punch, W. F. (2003, November). Predicting student performance: an application of data mining methods with an educational web-based system. In *33rd Annual Frontiers in Education, 2003. FIE 2003.* (Vol. 1, pp. T2A-13). IEEE.
[23] Ministrstvo za vzgojo in izobraževanje. MVI (2024A). Odredba o merilih za izbiro kandidatov v primeru omejitve vpisa v programe srednjega poklicnega izobraževanja, srednjega strokovnega izobraževanja ter gimnazij. *Uradni list republike Slovenije. (Official Gazette of the Republic of Slovenia).* https://pisrs.si/pregledPredpisa?id=SKLE13777
[24] Ministrstvo za vzgojo in izobraževanje. MVI. (2024B). Pravilnik o nacionalnem preverjanju znanja v osnovni šoli. *Uradni list republike Slovenije. (Official Gazette of the Republic of Slovenia).* https://www.uradni-list.si/glasilo-uradni-list-rs/vsebina/2024-01-2209/pravilnik-o-nacionalnem-preverjanju-znanja-v-osnovni-soli
[25] Ministrstvo za vzgojo in izobraževanje. MVI. (2024C). *Nova merila za izbiro kandidatov v primeru omejitve vpisa v srednješolske programe*. https://www.gov.si/novice/2024-08-23-nova-merila-za-izbiro-kandidatov-v-primeru-omejitve-vpisa-v-srednjesolske-programe/
[26] Namoun, A., & Alshanqiti, A. (2020). Predicting student performance using data mining and learning analytics techniques: A systematic literature review. *Applied Sciences*, *11*(1), 237.
[27] Okubo, F., Yamashita, T., Shimada, A., and Ogata, H. (2017). A neural network approach for students' performance prediction. In: *Proceedings of the Seventh International Learning Analytics & Knowledge Conference,* p. 598-599.
[28] Oloruntoba, S. A., and Akinode, J. L. (2017). Student academic performance prediction using support vector machine. *International Journal of Engineering Sciences and Research Technology, 6*(12): 588–597, 2017.
[29] Pallathadka, H., Wenda, A., Ramirez-Asís, E., Asís-López, M., Flores-Albornoz, J., & Phasinam, K. (2023). Classification and prediction of student performance data using various machine learning algorithms. *Materials today: proceedings*, *80*, 3782-

3785.
[30] Powers, David M. W. (2011). "Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation". Journal of Machine Learning Technologies. 2 (1): 37–63.
[31] Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, *73*(2), 417-458.
[32] RTV SLO and STA. (2024). Začenja se novo šolsko leto. Kakšne novosti prinaša? https://www.rtvslo.si/slovenija/zacenja-se-novo-solsko-leto-kaksne-novosti-prinasa/719477
[33] Sanders, W. L., & Rivers, J. C. (1996). *Cumulative and residual effects of teachers on future student academic achievement.* Research progress Report. University of Tennessee.
[34] Shabani, K., Khatib, M., & Ebadi, S. (2010). Vygotsky's zone of proximal development: Instructional implications and teachers' professional development. *English language teaching*, *3*(4), 237-248.
[35] Shahzad, K., & Naureen, S. (2017). Impact of Teacher Self-Efficacy on Secondary School Students' Academic Achievement. *Journal of Education and Educational Development*, *4*(1), 48-72.
[36] Shaziya, H., Zaheer, R., and Kavitha, G. (2015). Prediction of students performance in semester exams using a Naïve Bayes classifier. *International Journal of Innovative Research in Science, Engineering and Technology, 4*(10), 9823–9829.
[37] SIOL and STA (2024). Omejitve vpisa v srednje šole: ključno bo lahko tudi nacionalno preverjanje znanja. https://siol.net/novice/slovenija/omejitve-vpisa-v-srednje-sole-kljucno-bo-lahko-tudi-nacionalno-preverjanje-znanja-642535
[38] Susnjak, T. (2024). Beyond predictive learning analytics modelling and onto explainable artificial intelligence with prescriptive analytics and ChatGPT. *International Journal of Artificial Intelligence in Education*, *34*(2), 452-482.
[39] Tempelaar, D., Rienties, B., & Nguyen, Q. (2020). Subjective data, objective data and the role of bias in predictive modelling: Lessons from a dispositional learning analytics application. *PloS one*, *15*(6), e0233977.
[40] Tharwat A. (2018). Classification assessment methods. *Applied Computing and Informatics,* 17, 168–192.
[41] Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The relation between self-beliefs and academic achievement: A meta-analytic review. *Educational psychologist*, *39*(2), 111-133.
[42] Vygotsky, L. S. (1987). Thinking and speech (N. Minick, Trans.). In R. W. Rieber & A. S. Carton (Eds.), The collected works of L. S. Vygotsky. Vol. 1. *Problems of general psychology* (pp. 39–285). New York: Plenum Press
[43] Wongvorachan, T., He, S., & Bulut, O. (2023). A comparison of undersampling, oversampling, and SMOTE methods for dealing with imbalanced classification in educational data mining. *Information*, *14*(1), 54.
[44] Yağcı, M. (2022). Educational data mining: prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, *9*(1), 11.
[45] Yu, X., Wang, X., Zheng, H., Zhen, X., Shao, M., Wang, H., & Zhou, X. (2023). Academic achievement is more closely associated with student-peer relationships than with student-parent relationships or student-teacher relationships. *Frontiers in psychology*, *14*, 1012701. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9978389/

[46] Zimmerman, D. J. (2003). Peer effects in academic outcomes: Evidence from a natural experiment. *Review of Economics and statistics*, *85*(1), 9-23. https://www.jstor.org/stable/3211619

Acquiring competencies for mastering tools that process big data - the case of AI usage skill gaps

Alan Berg, Stefan Mol
Universitein van Amsterdam
1012 WP Amsterdam, The Netherlands
{a.m.berg, s.t.mol}@uva.nl

Scott Harrison, Gábor Kismihók
Technishe Informationsbibliothek
Welfengarten 1B, 30167 Hannover, Germany
{scott.harrison, gabor.kismihok}@tib.eu

Alenka Pandiloska Jurak
Faculty of Information Studies Novo mesto
Ljubljanska cesta 31A, 8000 Novo mesto, Slovenia
alenka.pandiloska@fis.unm.si

Janez Kolar
Science and Research Center RUDOLFOVO Novo mesto
Podbreznik 15, 8000 Novo mesto, Slovenia
janez.kolar@rudolfovo.eu

Leo Mršić
Algebra University
Gradišćanska 24, 10000 Zagreb, Croatia
leo.mrsic@algebra.hr

Abstract:
This paper delves into the critical role of Artificial Intelligence (AI) in shaping the 21st century, emphasizing its potential to revolutionize various sectors including medicine and education, while also posing significant societal and economic challenges. Amidst the backdrop of the EU's lag in AI adoption compared to the USA and Asia, this study underscores the acute skills gap as a pivotal hindrance. It introduces the AI4VET4AI project aimed at bridging this gap through a transnational European platform of Vocational Education and Training (VET) excellence. The project seeks to integrate AI-related content into VET curricula, fostering trust in AI, and aligning skills development with the Digital Education Action Plan (2021-2027). By highlighting the importance of a collaborative knowledge triangle—business, education, and research—the paper advocates for a comprehensive approach to harness AI's potential, ensuring sustainable growth and competitiveness within the EU.

Keywords: *Artificial Intelligence, skills gap, VET*

1. Introduction

The significance of Artificial Intelligence in the 21st century cannot be overstated. Experts predict that within just two decades, the impact of AI will be so profound that aspects of daily life will undergo a drastic transformation. The influence of AI is poised to challenge the very foundations of our economic and social systems. While it has the potential to create unprecedented wealth and revolutionize fields such as medicine and education, it also poses existential risks to our way of life [1]. This emphasizes the importance of the EC2020 report, which highlights the EU's concerning lag in AI adoption and development compared to the USA and Asia. A notable factor behind this discrepancy is the shortage of skilled workforce [2], [3].

The European Commission's Expert Group on AI, the European AI Alliance [4], and the coordinated plan with member states all emphasize the importance of consolidating European efforts in the AI strategy. The European Commission believes that only by doing so can the full potential of AI be harnessed to keep pace with global technological progress. Despite significant efforts, the EC's 2020 report reveals that Europe lags behind the USA and Asia in technology and AI. Eurostat data from 2020 shows that only 7% of EU enterprises with at least 10 employees used AI applications [5]. In 2023, 8% of EU enterprises uses the artificial intelligence technologies [6]. Additionally, 2% of enterprises utilized machine learning for internal big data analysis, while 1% employed natural language processing, natural language generation, or speech recognition for the same purpose. Only 2% of enterprises used chat services with chatbots or virtual agents generating natural language responses for customers. Similarly, 2% of enterprises utilized service robots with various degrees of autonomy for tasks such as cleaning, handling dangerous substances, sorting items in warehouses, and assisting customers at points of sale [5].

2. Addressing the skills gap in AI usage

In the contemporary business environment, significant generally acknowledged gap of skilled professional, perhaps a short snapshot from the past could be quite beneficial. Authors [7] are writing in this context also about Person-Organisation fit (P-O fit) concept which is gaining meaning due to changing organizational environments. In order for a candidate, which might also just came out the VET education, to fit the organization and optimally exploit his/her potentials, they identified five groups (clusters) of competencies, as follows: 1.Social competencies (Social competence, also referred to as interpersonal competence, that are defined as the ability to engage in cooperative interaction and communication with others. This cluster of competencies encompasses understanding others' thoughts, emotions, and attitudes, and therefore demonstrating empathy towards them. Recruiting individuals with strong interpersonal skills can be highly beneficial, especially in entrepreneurial contexts, as it facilitates the formation of new communication networks and access to critical resources); 2.Personal competencies (Personal competence is defined as the individual’s potential to adjust and change one's emotions, motivations, attitudes, and values in response to continuously challenging circumstances. It also refers to the ability to extend personal capabilities toward ethical actions, cultivating a self-image.); 3.Subject competence (often named in the literature as business, technical, or operational competence, are developed through professional training and experience. This cluster involves not only having technical expertise but also the ability to apply this knowledge effectively and appropriately in a timely manner); 4.Method Competencies cluster (which is defined as the ability to approach and address

challenges through structured problem-solving techniques. From the other point of view, it does involve the effective application of methods to efficiently plan, prepare, organize, use time as a resources, and use knowledge appropriately. This competency cluster clearly emphasizes the systematic use of appropriate strategies to navigate tasks and achieve desired outcomes.); 5.Entrepreneurial competencies cluster (Carrying out tasks efficiently, combined with the ability to anticipate and strategically manage decisions through a balance of intuition and foresight, is essential for evaluating the costs and benefits of every choice. Entrepreneurial competence is crucial for company founders to ensure long-term success. Additionally, the concept of corporate entrepreneurship (internal entrepreneurship, spin-offs, spin-outs and other forms of corporate entrepreneurship encouragements are highlighted in the literature, which encourages employees to focus on achieving organizational objectives, learning from setbacks, and safeguarding company resources. This approach combines the entrepreneurial capabilities of individuals to the overall success of the organization.), which, in the right environments, can bring us back to the first cluster of social competencies.

Within the project AI4VET4AI, a transnational European platform of VET excellence will be established to address the current skills gap in AI. The introduction of AI has significant implications for the work and VET, requiring VET providers to prepare students for new roles and tasks, handle increased work complexity, and adapt to new forms of collaborative work with AI. To confront these changes, VET providers need to embrace digital transformation efforts.

The leverage of the triangle of knowledge (business, education, and research) will be used to maintain high standards of VET excellence while integrating AI-related content into existing VET curricula or developing new ones. The overarching goals include fostering trust in AI, promoting understanding and awareness of reliable AI development and use in accordance with EU and international standards, and introducing AI-based innovation in key sectors through skills development aligned with the Digital Education Action Plan (2021-2027) [8]. The project aims to contribute to European education and training policy, as well as the development and innovation agenda, by emphasizing intelligent, enduring, and comprehensive growth in alignment with the Europe 2020 strategy. This strategy seeks to enhance EU competitiveness and growth by prioritizing economic transitions and ambitious skills development targets. The imperative to enhance the exploration and utilization of AI, as well as the concerns regarding how AI advancements can aid in addressing critical global challenges like climate change, are internationally significant.

The field of research in this area is complex, so we limited ourselves to the following research questions:

RQ1: What are the current skill gaps at the intersection of Al related skills and VET?

RQ2: What Al skills are in increasing demand from 2020 onwards, for VET students or graduates in the labour marketplace?

RQ3: Which occupations are going to need an Al skilled, VET qualified workforce, the most in the next 10 years?

RQ4: What skills belong to the Al literacy skill set?

RQ5: Are there any country or regional differences in the Al skills gap by occupation?

RQ6: What methods can be implemented to provide ongoing and adaptive recommendations for the development of state-of-the-art Al skills curricula?

RQ7: What are the most effective methods for disseminating Al teaching materials?

To provide a focus we refined the Research Questions into three themes:

1. Al skills gap analysis: This theme is related to RQ1, RQ2, RQ3, and RQ5. These

questions aim to identify the current and future skill gaps at the intersection of Al and VET and how they vary by occupation, country, or region. The theme can help understand the demand and supply of Al skills in the labour market and the challenges and opportunities for VET students or graduates.
2. Al literacy skill set: This theme is related to RQ4. This question aims to define the skills of the Al literacy skill set, which are essential for understanding, using, and interacting with Al systems and applications. This theme helps establish a common framework and standard for Al education and training.
3. Al teaching methods: This theme is related to RQ7. This question explores the most effective methods for disseminating Al teaching materials, such as on line platforms, courses, modules, or workshops. This theme can help to improve the accessibility and quality of Al learning resources and experiences.

The basis for being able to reach the intended goal was to identify the specific needs of important industries and sectors across 10 countries and 17 NUTS2 regions represented in the consortium, in line with their existing S3 frameworks. Further on, a series of MOOCs (Massive Open Online Courses) will be developed, targeting the benefits and challenges of AI in the most promising industry sectors identified regionally. Additionally, support for VET teacher training will be provided.

3. Research methodology

The research operationalization has been prepared by the team at the University of Amsterdam and Technishe Informationsbibliothek. The study was initiated through a collaborative partnership of 26 partners from 11 different countries (Croatia, Slovenia, Serbia, Austria, Germany, Greece, Italy, Spain, Netherlands, Turkey, and Ireland). The cross-border approach helped to mitigate the potential problem of the limited availability of AI-related training opportunities in underdeveloped regions, which could exacerbate the gap in innovation and development between more and less developed areas of Europe and hinder their transition to an AI-driven economy and its associated sustainable development objectives. The AI4VET4AI project assesses the influence of Artificial Intelligence on Al literacy in European Vocational Education and provides recommendations. This is achieved by conducting a skills gap analysis focusing on sector and regional variations. The effects and speed of Al on different demographics, regions, industries, and cultures are intricate and may only possibly be fully comprehended by future historians. As a result, the authors utilized multiple methodologies to corroborate their findings, employing a mixed methods approach to produce recommendations for future curricula accordingly.

The research methodology set included:
1. Skills gap survey
2. Delphi method. A discussion with experts
3. Systematic Literature Review of skills demand in the job market via Job market intelligence that, in the end, requires the analysis of a big set of on line Job advertisements
5. Systematic data review of available sources of data. Later, a number of the discovered data sources are used to support the recommendations.

The research utilized rigorous methods to establish well-informed recommendations supported by multiple sources. We made specific recommendations for AI literacy by conducting a systematic literature review, whereby ten researchers initially identified a large number of research papers and narrowed it down to 102 high-quality papers specifically related to the theme of AI literacy. One limitation we encountered was the freshness and quality of research and recommendations, which are impacted by

intense competition in advancing AI architecture and the rapid evolution of AI models trained on vast amounts of internet data. The widespread adoption of AI is rapidly affecting various industries, with implications for workers as the pace of change in these industries outstrips others.

To address the challenge of providing timely advice, we launched a systematic data review. However, the velocity of change in the field of AI limits the quality and freshness of datasets and information related to AI literacy. Adhering to a "Chain of Trust" methodology, we initially selected papers that had undergone peer review and trusted the quality of the sources cited within. Among the most valuable resources have been observatories established mainly by governance or standards bodies to monitor specific AI trends. Additionally, relevant reports generated periodically by international organizations have provided valuable insights and data sources for policymakers.

In our review, we primarily relied on Eurostat, the statistical office of the European Union, for regional data. However, we encountered limitations related to the freshness and geographical resolution of the datasets, which affected the recommendations. Our focus was on specific industries and regions, aligning with current economic data and the capacity of the Al4VET4AI consortium. The industries and regions were selected based on input from the 27 project partners and a survey of experts.

Through our methodology, we conducted a comprehensive analysis of the AI skills landscape in Europe and identified gaps between the supply and demand of AI skills, as well as best practices and recommendations for addressing them. In the next chapter, we will provide the results of the systematic literature review, offering an overview of the current state of the art in AI skills research and development.

4. First results

Based on the results of research, implemented as described in chapter 3, five sectors were selected for further analysis: (i) Professional, Scientific and Technical Activities, (ii) Wholesale and Retail Trade, (iii) Human Health and Social Work, (iv) Education and Accommodation and (v) Food Services. Further analysis was set to find converging evidence of key skills gaps. Based on the results, an annotated list of skills was prepared.

In the next step, all previous results are taken into account and the first drafts of IVET and CVET syllabuses are being created as follows:

- Accommodation and Food Services

(i) Customer Relationship Management - using AI to provide better service to customers
(ii) Using AI for Understanding Tourism Impact and Improving Sustainability

- Education

(i) Using AI for process automation in teaching and for creating personalised recommendations for learner
(ii) Human-AI Collaboration in Teaching and Learning

- Human Health and Social Work

(i) Case Study: Synthetic Patient Records how and why
(ii) Supporting Doctor - Patient Interactions with AI

- Professional, Scientific and Technical Activities (Accounting, Bookkeeping and auditing activities; tax consultancy / Advertising / Market Research and public opinion polling)

(i) Using AI for personalised advertising campaings
(ii) AI in market research (sentiment analysis, machine learning for trend prediction, ethics and privacy in AI-driven market research)

- Wholesale and Retail Trade

(i) AI in Customer Service - use of chatbots / virtual assistants and personalised suggestions
(ii) AI-Driven Inventory and Logistics Automation

All other data and obtained results are marked as internal until the end of the project, so data sharing is not allowed.

5. Where do we go from here

Based on the findings, the project will facilitate the creation of VET innovation hubs, where VET learners can acquire skills in using AI to enhance their problem-solving abilities. As AI's introduction to the labour market emphasizes the importance of VET learners' social competences and personal skills, our project addresses a range of target groups.

• VET learners (developing AI-related competences and skills)
• VET teachers (including teaching of AI-related content in their courses and broadening the scope of their teaching)
• HEI students and teachers (forming partnerships with VET organisations for the purpose of exchanging knowledge and innovation practices)
• sectoral representatives (gaining partners who will contribute to closing AI skills gap)
• Innovative startups, SMEs and other business organisations (acquiring the much needed talent pool of experts)
• Policy makers (on regional and national level, including them in a more focused way on how VET Centres of Excellence can contribute to regional smart specialisation strategies)

The goal of AI4VET4AI is to contribute to the digital transformation of the EU labour market by introducing new innovative teaching content and methods to VET curricula in 11 European countries and 18 EU NUTS2 regions. This initiative aims to support the development of AI-skilled workers. Starting with a comprehensive analysis of the most promising sectors for AI deployment in our 17 regions, we collaborate closely with enterprises and their cluster organizations to develop 14 MOOCs and TT materials that can be easily integrated into VET programs (IVET and CVET). Additionally, we are establishing 11 innovative AI VET campuses and 7 VET innovation AI incubators to provide VET learners with opportunities to enhance their creativity and entrepreneurial skills. Through project activities, we foster close collaboration among partners and raise awareness of the potential of AI in our regions among stakeholders from the public and private sectors, as well as civil society. This enables us to build a collective and active platform comprising EU citizens and institutions who are interested, well-informed, and motivated to support further AI development. Ultimately, this forms the basis of our ambitious CoVE (Centre of vocational excellence), which aims to attract a wide array of institutions (HEIs, VETs, companies, agencies, and individuals) in its quest to create a sustainable, inclusive, and equitable AI-powered future for all.

6. References:

[1] ‘AI4VET4AI project’, AI4Vet4AI. Accessed: Sep. 16, 2024. [Online]. Available: https://www.ai4vet4ai.eu/
[2] European Commission, ‘Artificial Intelligence: first quantitative study of its kind finds uptake by businesses across Europe is on the rise | Shaping Europe’s digital future’. Accessed: Sep. 16, 2024. [Online]. Available: https://digital-strategy.ec.europa.eu/en/news/artificial-intelligence-first-quantitative-study-its-

kind-finds-uptake-businesses-across-europe
[3] European Commission, ‘White Paper on Artificial Intelligence – a European approach to excellence and trust | Shaping Europe’s digital future’. Accessed: Sep. 16, 2024. [Online]. Available: https://digital-strategy.ec.europa.eu/en/consultations/white-paper-artificial-intelligence-european-approach-excellence-and-trust
[4] European Commission, ‘The European AI Alliance | Shaping Europe’s digital future’. Accessed: Sep. 17, 2024. [Online]. Available: https://digital-strategy.ec.europa.eu/en/policies/european-ai-alliance
[5] Eurostat, ‘Artificial intelligence in EU enterprises’. Accessed: Sep. 16, 2024. [Online]. Available: https://ec.europa.eu/eurostat/web/products-eurostat-news/-/ddn-20210413-1
[6] Eurostat, ‘Use of artificial intelligence in enterprises’. Accessed: Sep. 16, 2024. [Online]. Available: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Use_of_artificial_intelligence_in_enterprises
[7] Zu Knyphausen-Aufseß, D., & Vormann, C. (2009). Personnel selection criteria in IT ventures: A policy-capturing analysis. Zeitschrift Für Betriebswirtschaft, 79(2), 213–234. doi:10.1007/s11573-008-0207-8
[8] European Commission, ‘Digital Education Action Plan (2021-2027) - European Education Area’. Accessed: Sep. 17, 2024. [Online]. Available: https://education.ec.europa.eu/focus-topics/digital-education/action-plan

Acknowledgements

The research is co-financed by the European Union, European Education and Culture Executive Agency; AI-powered Next Generation of VET, 101104579 — AI4VET4AI — ERASMUS-EDU-2022-PEX-COVE. The researchers from FIŠ and Rudolfovo are partly funded by the Slovenian Research and Innovation Agency (ARIS) through the annual work programme of Rudolfovo and by the Slovenian Ministry of Higher Education, Science and Innovation.

Next to the TiB and UvA project teams, and above mentioned co-authors, we gratefully acknowledge Serdjan Verbić, Stephan Schloegl, Willemijn van Kooten, Monika Rohlik Tunjić, Maja Brkljačić, Annika Gnädinger, and Silvia Cosimato for their contributions to conceptualization, screening, and coding.

Intellectual Property Rights Data for Sustainable Digitalisation and Digitalisation for Sustainability

Ana Hafner
Rudolfovo – Science and Technology Centre Novo mesto
Podbreznik 15, 8000 Novo mesto, Slovenia and
Faculty of Information Studies
Ljubljanska cesta 31A, 8000 Novo mesto, Slovenia
ana.hafner@rudolfovo.eu

Dolores Modic
Nord University Business School, Nord University
Postboks 1490, 8049 Bodø, Norway and
Rudolfovo – Science and Technology Centre Novo mesto
Podbreznik 15, 8000 Novo mesto, Slovenia
dolores.modic@nord.no

Abstract: *This paper explores the relationship between sustainability, digitalisation, and intellectual property rights, by focusing on the role of the IP Monitor web portal. While IP Monitor provides a one-stop-shop for intellectual property rights data, and offers direct links to over 600 databases, as well as provides informational and educational content, its contribution lies also in promoting intellectual property for sustainability. The content analysis highlights the portal's alignment with EU sustainability goals but also notes its limitations, such as a lack of interactivity and critical assessments of how to effectively categorise sustainable inventions. Ultimately, the paper calls for improvements in user engagement and IP classification systems to enhance the platform's effectiveness in supporting sustainable practices.*

Key Words*: Intellectual property, big data, IP Monitor, sustainability, circular economy, digitalisation*

1 Introduction

Perceptions of whether digitalisation promotes sustainable development vary widely. On one hand, many see it as a tool for protecting the climate and natural resources, fostering development, and supporting a collaborative economy, on the other hand, some view digitalisation as a vehicle for increasing capitalistic control and surveillance over nature [1]. Some researchers claim that under the current political and economic system, digitalisation may exacerbate the ecological harm of industrialisation rather than alleviate it [2]. Concerns are connected with the extraction of minerals and increasing energy consumption of information and communication technologies, with operating connecting data facilities consume a significant amount of energy [3,4].

The terms "sustainable digitalisation" and "digitalisation for sustainability" both relate to the intersection of digital technology and sustainability, but they approach the relationship

from different angles [5]. Sustainable digitalisation is about making the digital technologies themselves sustainable. Digitalisation for sustainability is about using digital technologies to make other processes and sectors more sustainable.

Data within intellectual property rights (IPR) (e.g. patent, trademark) databases represent a vast collection of big data [6], [7] encompassing millions of entries from global patent offices and trademark registries. This wealth of information holds immense value for innovation, business strategy, and sustainability efforts. Data is the oil of the digital era, but as Lucivero [8] noticed, just like oil, data production, collection and processing have an environmental impact.

The aim of this paper is to explore what are the connections between "sustainability", "digitalisation", and "IPR" and how IPR can contribute to "sustainable digitalisation" and "digitalisation for sustainability" whilst focusing on the IP Monitor solution developed within the IPR4SC project.

2 Method

This paper will present a case study of the web portal IP Monitor [9], a platform designed to track intellectual property developments, also with a particular emphasis on sustainability and circular economy. This case study approach includes a qualitative content analysis of data and content from IP Monitor. Through this, the study aims to understand how IPR can influence both sustainable digitalisation and digitalisation for sustainability.

The selection of IP Monitor as a case study is driven by its unique focus on the intersection of intellectual property and sustainability and circular economy. The selection process of the tools included in the IP monitor was multi-step.

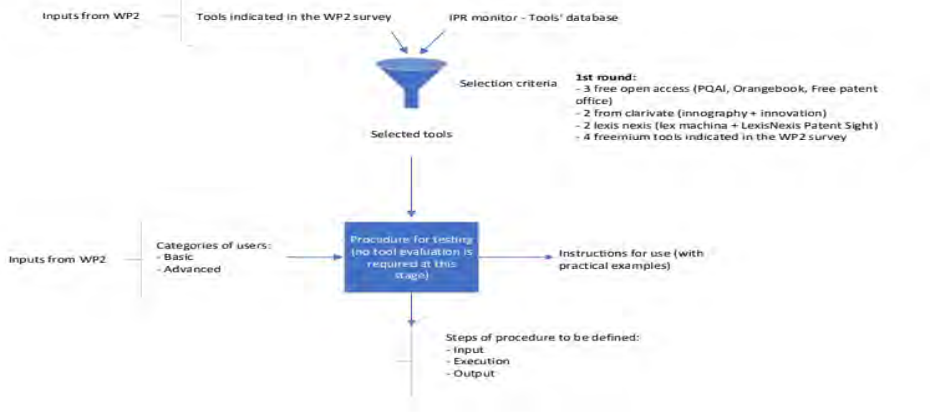


Figure 1: Protocol for analysis of IP tools in IPR4SC (WP3)

As initial steps preliminary inputs were provided from the previous IPR4SC project survey, where we focused especially on the accessibility of the tools (see Figure above). Next, testing procedure was developed and an initial screening was done.

Lastly, the tools underwent a more in-depth analysis. For this, we constructed two cases, where one was constructed solely from prompts to ChatGPT (except for the related Figure), and the other one based on a real-life case. Case 1 was generated with the help of ChatGPT in order to avoid similarities with the existing patent text as well as minimising any potential secrecy issues that might ensue by using a real-live case. Furthermore, this also allowed for an insight on the usability of the ChatGPT in this kind of exercises. The analysis focused on functionalities related to search engine functions, analytics functions and docketing functions. However, more specifically we also focused on two issues one related to AI functionalities of the tools, and the other was related to functionality related to green, circular and sustainable IP – which is of particular interest to the topic at hand. By analysing the portal's role in fostering or hindering sustainable practices, this study seeks to uncover how IP tools can support both environmental and technological advancements.

3 Findings

IP Monitor is a web portal which educates about IP (General information) [10], but this is not its main mission. IP Monitor also presents a collection of IPR databases, practically all databases that currently exist worldwide (IP Monitor database) and contains over 600 databases and other resources [11]. They are sorted into four sections: 1) Databases, 2) IP tools, 3) Technology transfer opportunities and 4) Registries of IP attorneys. Therewith the IP Monitor presents the world's largest IP portal to access a range of resources from one place. But how is this connected to sustainable digitalisation? Offering a centralised platform represents time savings and energy savings when the user is searching the internet how to identify the most useful IP tool or database for the issue at hand. This enables businesses, researchers, and policymakers to track IPR data in one place and quickly discover what they are searching for. This can result in savings in time and electricity costs connected with the use of information technology, directly contributing to sustainable digitalisation by optimising resource usage. In other words, big data produces big waste [8], and any simplification and optimisation on how to access big data can reduce this waste.

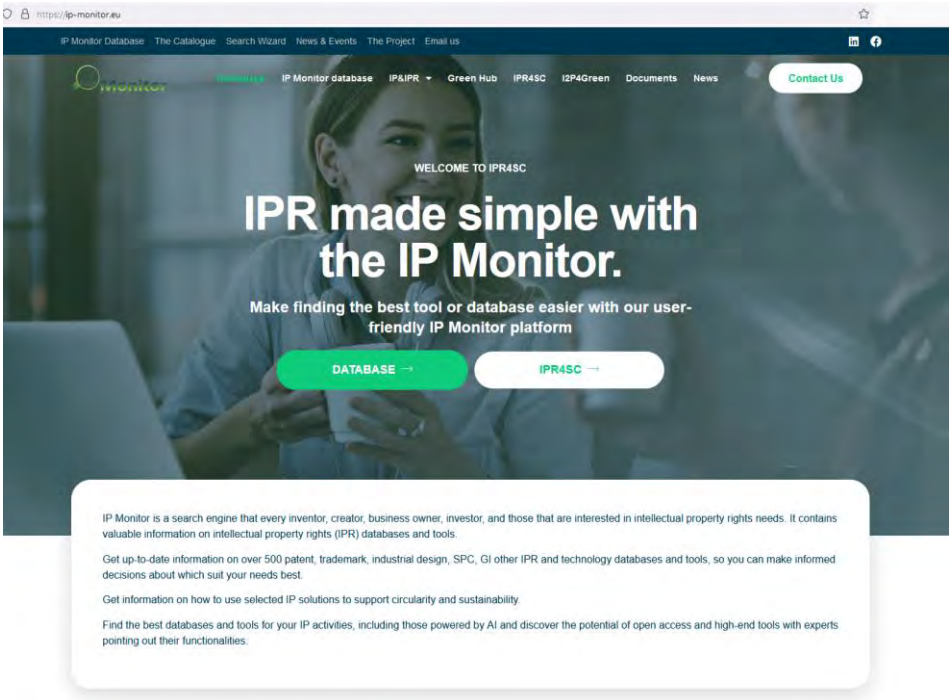


Figure 2: IP Monitor – title page

However, IP Monitor's greater contribution lies in digitalisation for sustainability. By promoting sustainability and circular economy with Green Hub [12] and I2P4Green course [13] and by providing access to a wide range of technology transfer opportunities, IP Monitor plays an important role in accelerating the adoption of sustainable technologies. Businesses and innovators can search for green technologies and intellectual property that can drive sustainability across sectors. This promotes the dissemination of eco-friendly innovations, enabling industries to integrate sustainable solutions faster and more efficiently.

One of the aims of the IP Monitor was also to tease out solutions in terms of IP databases and IP tools that allow for functionality related to green, circular and sustainable IP, or to identify and highlight those that are specifically focused on them. In terms of the latter, some can be mentioned. IP Monitor's Green Hub [12] especially emphasises WIPO Green database [14] and other IP initiatives such as Green EU Trademarks [15] and EU Ecolabels [16]. In this way, IP Monitor facilitates the alignment of technological advancements with global and EU sustainability goals. However, the IP Monitor had not been able to provide insights on the particular functionalities related to green, other than some tools allowing to focus on the EPOs tagging in terms of climate change inventions.

There are also some other limitations we can identify. The portal lacks interactivity and does not invite users to participate with comments, for example, through online forums. This is relevant as much of the circularity and sustainability literature emphasises the role of the end-users to advance and promote circular and sustainable solutions [17], [18]. We also lack a critical assessment of what could be done at the IP data science level to make

green and ecological solutions as easy as possible to identify in databases. Improving these aspects could significantly boost the platform's effectiveness and user engagement in promoting sustainability.

4 Discussion and conclusion

IP Monitor is a web portal that serves as an educational resource on intellectual property (IP) and offers an extensive collection of over 600 IPR databases and tools, categorised into four sections. It supports sustainable digitalisation by enabling users to efficiently track IP data, thereby optimising resource usage and reducing duplication of efforts. Its greater contribution lies in promoting sustainability and circular economy (CE), as it facilitates access to green technologies and technology transfer opportunities. While it highlights initiatives like the WIPO Green database, the platform lacks interactivity, user engagement features and critical assessment in the field of how IP can contribute to sustainability.

For example, IP Monitor does not deal in-depth with Green Patent Classification systems, such as WIPO's Green Inventory, the OECD's ENV-TECH or the European Patent Office's (EPO) Y-tags. The Y-tagging schemes were developed by experts from within the EPO with the help of several external partners that also provided specialists to evaluate the technologies [19]. Y-tags are metadata tags used to categorise and identify IP assets, particularly in relation to environmental sustainability. This categorization supposedly helps to promote the visibility and accessibility of eco-friendly inventions, however, they are not without flaws. As Modic et al. noted [20], current indicators of CE-related (technical) innovations exclude patent documents we need for CE innovation management and public policy research questions. How to efficiently and uniformly identify sustainable inventions within patent databases, thus remain a challenge for the future.

IP Monitor could be more proactive in addressing gaps in the recognition and classification of sustainable technologies within patent databases. Developing a more comprehensive framework for identifying and tagging eco-friendly innovations will help businesses, researchers, and policymakers make better use of IP data for sustainability-driven decisions. By refining its categorization systems and fostering a collaborative environment for users, IP Monitor can further contribute to both sustainable digitalisation and digitalisation for sustainability.

Acknowledgement

Funding acknowledgement: This work has been co-funded by the Developing Skills in Intellectual Property Rights Open Data for Sustainability and Circularity (IPR4SC) project funded by the Erasmus + Alliances for Innovation programme, grant number 101056129-IPR4SC. Authors also gratefully acknowledge the support from the Slovenian Research and Innovation Agency (ARIS) through the annual work program of Rudolfovo and the KTO3 project, co-funded by Republic of Slovenia, Ministry of Higher Education, Science and Innovation and European Union.

References

[1] O. Enocksson, *The green and digital transition : A case study about the EU twin strategy and the perceptions of six large companies in Sweden*. 2023. Accessed: Sep. 29, 2024. [Online]. Available: https://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-334184

[2] S. Kunkel and D. Tyfield, 'Digitalisation, sustainable industrialisation and digital rebound – Asking the right questions for a strategic research agenda', *Energy Res. Soc. Sci.*, vol. 82, p. 102295, Dec. 2021, doi: 10.1016/j.erss.2021.102295.

[3] J. Arribas Cámara, 'Humanity Is Facing Its Sustainability: Will Technological Progress Make the Future Unsustainable?', in *Science, Technology and Innovation in the History of Economic Thought*, E. Trincado Aznar and F. López Castellano, Eds., Cham: Springer Nature Switzerland, 2023, pp. 241–256. doi: 10.1007/978-3-031-40139-8_12.

[4] M. A. B. Siddik, A. Shehabi, and L. Marston, 'The environmental footprint of data centers in the United States', *Environ. Res. Lett.*, vol. 16, no. 6, p. 064017, May 2021, doi: 10.1088/1748-9326/abfba1.

[5] P. Hentunen, 'The Impact of Digitalisation on Sustainability: Case Study of Renewable Materials Company'.

[6] N. Damij, A. Hafner, and D. Modic, 'Activity-to-Skills Framework in the Intellectual Property Big Data Era', *IEEE Trans. Eng. Manag.*, vol. 71, pp. 13251–13265, 2024, doi: 10.1109/TEM.2022.3175738.

[7] D. Modic, A. Hafner, N. Damij, and L. Cehovin Zajc, 'Innovations in intellectual property rights management: Their potential benefits and limitations', *Eur. J. Manag. Bus. Econ.*, vol. 28, no. 2, pp. 189–203, Jul. 2019, doi: 10.1108/EJMBE-12-2018-0139.

[8] F. Lucivero, 'Big Data, Big Waste? A Reflection on the Environmental Sustainability of Big Data Initiatives', *Sci. Eng. Ethics*, vol. 26, no. 2, pp. 1009–1030, Apr. 2020, doi: 10.1007/s11948-019-00171-7.

[9] 'IPR4SC - IPR made simple with the IP Monitor.', IPR4SC. Accessed: Sep. 29, 2024. [Online]. Available: https://ip-monitor.eu/

[10] 'General Information', IPR4SC. Accessed: Sep. 29, 2024. [Online]. Available: https://ip-monitor.eu/general-information/

[11] 'IP Monitor'. Accessed: Sep. 29, 2024. [Online]. Available: https://ip-monitor.eu/database/#step-1

[12] 'Green Hub', IPR4SC. Accessed: Sep. 29, 2024. [Online]. Available: https://ip-monitor.eu/green-hub/

[13] 'IPR4SC - I2P4Green course'. Accessed: Sep. 29, 2024. [Online]. Available: https://www.ipr4sc.eu/I2P4Green.html

[14] 'Wipogreen Database'. Accessed: Sep. 30, 2024. [Online]. Available: https://wipogreen.wipo.int/wipogreen-database/database

[15] European Union Intellectual Property Office., *Green EU trade marks: analysis of goods and services specifications, 1996 2021 : 2022 update.* LU: Publications Office, 2023. Accessed: Sep. 30, 2024. [Online]. Available: https://data.europa.eu/doi/10.2814/706335

[16] 'EU Ecolabel - Home'. Accessed: Sep. 30, 2024. [Online]. Available: https://environment.ec.europa.eu/topics/circular-economy/eu-ecolabel_en

[17] 'Sustainable Circular Mobility: User-Integrated Innovation and Specifics of Electric Vehicle Owners'. Accessed: Oct. 24, 2024. [Online]. Available: https://www.mdpi.com/2071-1050/12/19/7900

[18] B. Hazen, I. Confente, D. Pellathy, and I. Russo, ‘From End-of-the-road to Critical Node: The Role of End-user “Consumers” in Shaping Circular Supply Chain Management’, in *Circular Economy Supply Chains: From Chains to Systems*, L. Bals, W. L. Tate, and L. M. Ellram, Eds., Emerald Publishing Limited, 2022, pp. 151–165. doi: 10.1108/978-1-83982-544-620221008.
[19] A. N. Tanner, L. Faria, M. A. Moro, E. Iversen, R. Østergaard, and E. K. Park, ‘Regional Distribution of Green Growth Patents in four Nordic Countries: Denmark, Finland, Norway and Sweden’.
[20] D. Modic, A. Johnson, and M. Vučkovič, ‘Towards measuring innovation for circular economy using patent data’, in *Research Handbook of Innovation for a Circular Economy*, Edward Elgar Publishing, 2021, pp. 265–276. Accessed: Sep. 30, 2024. [Online]. Available: https://www.elgaronline.com/edcollchap/edcoll/9781800373082/9781800373082.00032.xml

Cluster Analysis of Key Links Between Digital Transformation and Sustainable Macroeconomic Growth in EU Countries

Dorian Fildor
The Institute of Economics, Department for Regional Development
Trg J. F. Kennedyja 7, Zagreb, Croatia
dfildor@eizg.hr

Abstract: *This paper explores the relationships between digital transformation, economic performance, and environmental sustainability in the European Union (EU). By integrating key macroeconomic indicators, such as GDP per capita, unemployment rates, and digital engagement metrics like the Digital Economy and Society Index (DESI), the study employs cluster analysis and Principal Component Analysis (PCA) to uncover significant correlations and patterns among EU member states. Findings indicate a strong positive correlation between digital transformation and economic performance, with countries exhibiting higher GDP per capita also demonstrating greater internet usage and digital engagement. Conversely, the study reveals a complex interrelationship between unemployment rates and CO2 emissions, suggesting that economic activities associated with high CO2 emissions may lead to an increase in unemployment, the alternative of which would be investment in renewable energy sources and digital transformation to create new jobs.*

Key Words*: Digital transformation, economic growth, environmental sustainability, macroeconomic indicators, cluster analysis, Principal Component Analysis, GDP per capita, CO2 emissions*

1 Introduction

Digital transformation is inevitably important in today's fast-growing technological environment. However, the aspect of observing the impact of digital transformation on the state of the economy is missing. Citizens should be more satisfied when they have access to digital services [1], whether from the state or private companies. After the COVID-19 pandemic [2], most services were digitized in order to be available to users at all, because otherwise the delivery of such services would be denied and users would probably be largely dissatisfied.

2 Literature overview

Digital transformation includes the integration of digital technologies into various aspects of society and economy, reshaping how organizations operate and deliver value to customers [3]. A growing body of literature has explored the relationship between digital transformation and economic growth. For instance, digital technologies have the potential to enhance productivity and innovation [4], which are crucial drivers of economic growth. Their work highlights that while digital tools can lead to significant efficiency gains, disparities in access and adoption can

exacerbate economic inequalities across different regions and sectors.

In the context of the European Union (EU), various studies have documented the progress of digitalization and its economic implications. The European Commission (2020) emphasizes the importance of a robust digital economy for sustainable growth, outlining strategies to improve digital infrastructure and skills across member states. Additionally, the assumption is that digitization is the basic setting for ensuring sustainable growth [5], and that digital innovations themselves achieve long-term growth. However, despite these advancements, the exact connection has not yet been confirmed with certainty.

3 Methodology

The research process begins with selecting the necessary variables for constructing a comprehensive model that captures the relationship between digital transformation and key macroeconomic indicators. The first step involves identifying indicators that reflect the state of digitalization in society [6], such as the rate of internet users and the Digital Economy and Society Index (DESI), which measures the performance of EU countries in terms of digital competitiveness. Alongside these, main macroeconomic variables such as GDP per capita in Purchasing Power Standards (PPS) and the unemployment rate [7] are selected to represent the economic performance of the society. To extend the analysis towards sustainable growth, additional variables are incorporated to capture environmental and green development dimensions. Indicators such as CO2 emissions per capita and the share of renewable energy resources [8] in total energy consumption are used to represent the sustainability aspect of economic growth.

Given the diversity and varying nature of these variables [9], a two-stage cluster analysis is employed, complemented by Principal Component Analysis (PCA) to handle the complexity [10]. In the first stage, PCA is used to reduce dimensionality and identify the most relevant features while maintaining data variability. A preliminary cluster analysis then examines the correlation coefficients among the selected variables to uncover potential relationships and patterns. This step helps to discern clusters of variables that may exhibit significant associations, guiding the identification of key indicators for digitalization, macroeconomic performance, and sustainable growth.

Following the initial analysis, a second-stage cluster analysis is carried out to refine the model by excluding irrelevant or weakly correlated variables. The refined set of variables provides a more robust framework for understanding the connections between digital transformation and economic outcomes. To integrate the selected variables from the domains of digitalization and macroeconomics, a mapping of coefficients is used to visualize and assess the significance of their interactions. Through this approach, the study aims to uncover the dynamic relationships and provide insights into how digital factors contribute to sustainable economic growth in EU countries.

4 Results

Most of the data was collected through Eurostat and refers to the year 2023. Only the data on unemployment have already been adjusted for seasonality and are available as such on Eurostat. The DESI indices were taken from the website European Commission. Data on CO2 emissions were obtained from edgar.jrc.ec.europa.eu, with data from non-independent territories also included alongside data for home countries. Data on the share of renewable energy sources were collected from the Energy Institute - Statistical Review of World Energy portal.

Table 1 Data about clusters

Cluster	GDP per capita in PPS	Unemployment rate (%)	CO2 per capita (%)	Share of renewable energy sources (%)	Internet users (%)	DESI index
0	123	5.4	6.65	84.48	96.34	75.18
1	118	5.6	7.18	33.06	96.76	67.05
2	64	4.4	5.86	25.91	88.18	43.79
3	95	5.5	5.83	20.22	91.59	59.76
4	91	2.7	8.52	14.87	94.83	65.93

Table 1 shows a table where there are five clusters (0 to 4) based on different socioeconomic indicators. Cluster 0 stands out as the strongest with a GDP per capita of 123 PPS, an unemployment rate of 5.4%, CO2 emissions (6.65%) and a high share of renewable energy sources (84.48%). Cluster 0 has the highest DESI index, which indicates that the countries in that cluster have achieved high levels of digital transformation. Cluster 1 has similar indicators, while cluster 2 shows the lowest economic strength with a GDP of 64 PPS, but also the lowest unemployment rate (4.4%).

Table 1 shows that the first cluster is the one with the highest income, the highest DESI index, and the most Internet users, which indicates that it is obviously the country with the greatest economic prosperity. The combination of the given inputs suggests that if the country implements a digital transformation, the given investments will be returned in the form of an increase in GDP per capita and a decrease in unemployment. Investing in technological infrastructure means reducing costs and increasing efficiency in the future. Considering that by increasing efficiency, the worker can

produce/deliver more in the same unit of time, the income from the labor of the working individual increases, and this increase has been seen through GDP per capita.

Cluster 4, although it has the lowest unemployment (2.7%), and the lowest share of renewable sources (14.87%). This analysis points out interesting trade-offs between economic performance, environmental impact and digital inclusion, which can help formulate sustainable development policies.

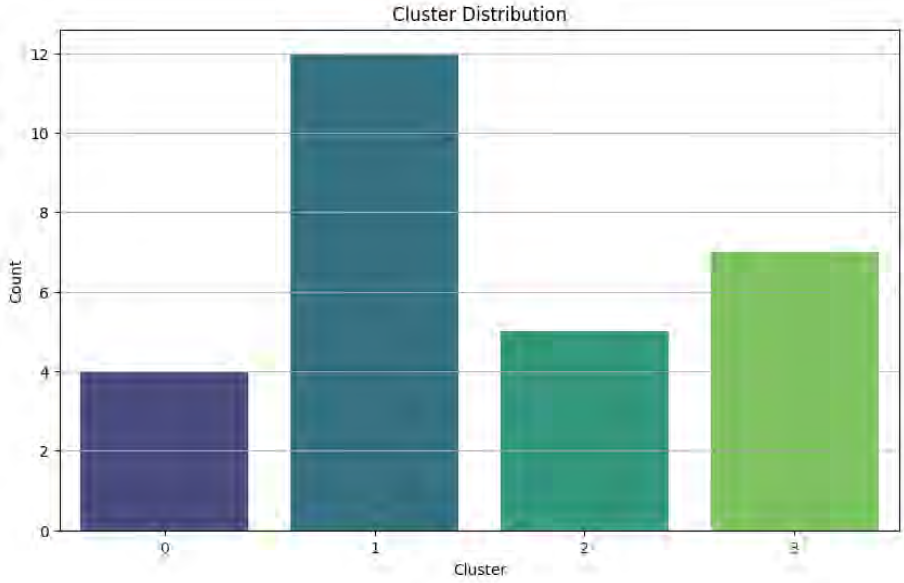


Figure 1 Distribution among clusters

Fig. 1 shows the distribution among the clusters, so the first cluster contains 4 countries, while the second cluster contains 12 countries and it is the largest cluster. The third cluster has 5 and the last cluster has 7 countries.

Fig. 2 shows the map of correlations between variables. Thus, PCA1 significantly correlates with indicators that indicate more successful economies. This is especially indicated by the correlation with GDP per capita in PPS with a coefficient of 0.86, which is a very strong connection. Significant correlations with the variables internet users with 0.81 and DESI index 0.77 are also in favor.

PCA2 significantly correlates with the unemployment rate 0.84. Also, the correlation between CO2 per capita and PCA2 is in a negative relationship and amounts to -0.64, which indicates that it is an economy that depends on the CO2 footprint. There is also a negative correlation between CO2 and the unemployment rate. On the other hand, CO2 and GDP per capita correlate 0.5.

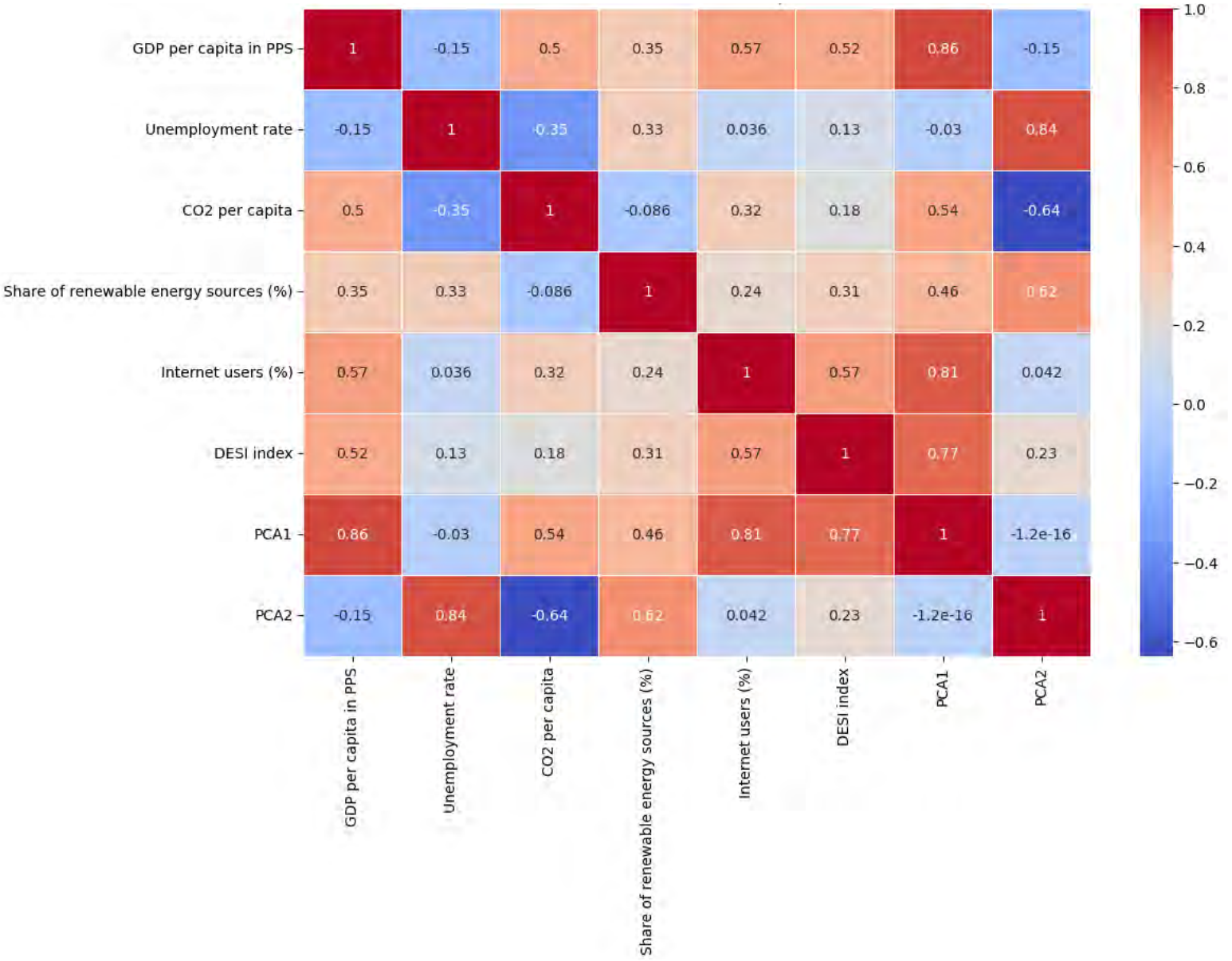


Figure 2 Correlation map

It is important to consider the context in which this correlation occurs. For example, in industries that are heavy polluters - probably traditional economies, reducing CO2 emissions can lead to reduced production and, potentially, increased unemployment. However, in other sectors, such as renewable energy or digital transformation, reducing CO2 emissions can create new jobs. Stopping economic activity in sectors of the traditional economy would probably direct workers and investments towards green technologies, which would increase unemployment in the short-term, which would return to the natural rate of unemployment in the long-term.

GDP per capita correlates with PCA1 with a coefficient of 0.86, while Internet users correlate with PCA1 in the amount of 0.81. The unemployment rate correlates with 0.84 on PCA2. All coefficients exceed 0.80, which is very close to the upper limit of positive correlation, therefore precisely those variables that have the greatest impact on PCAs were selected from the second step of cluster analysis in order to get a more detailed insight into the data with those variables included in the model.

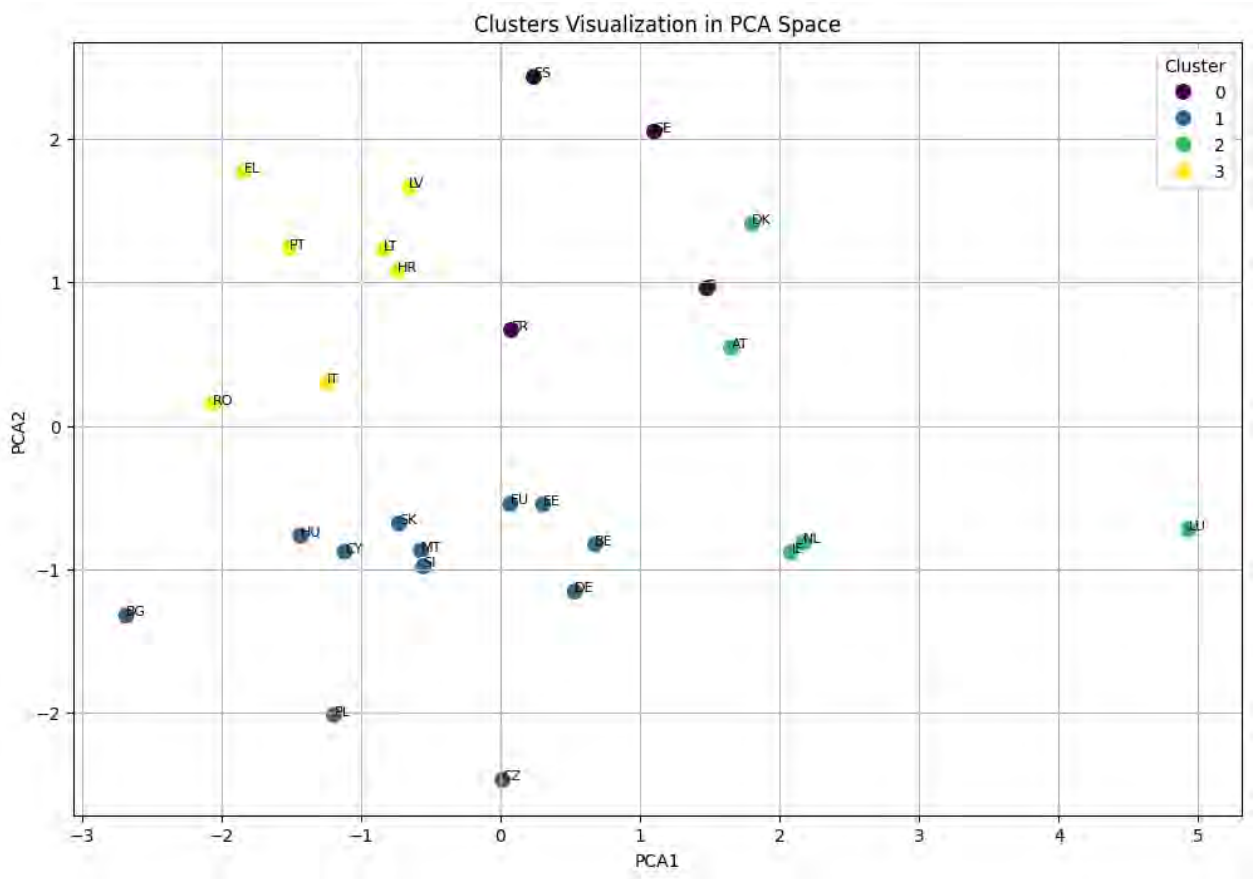


Figure 3 Clusters in PCA space

Fig. 3 shows the distribution of EU countries according to PCA1 and PCA2. It is obvious that PCA1 represents countries that are economically much stronger than the countries located on the ordinate symbolized by PCA2.

The complex relationship between socioeconomic indicators and sustainability in EU countries is analyzed through four figures. Table 1 indicates the differences between clusters, with cluster 0 not only showing the strongest economic indicators, but also high integration of renewable energy sources with very good other indicators, which can be a model for future development. Table 1 reveals that the largest cluster (Cluster 2) includes numerous countries with lower economic performance, suggesting that there is a need for targeted policies to empower these countries and

improve their positions.

Fig. 2 explores the correlations among the variables, highlighting that high CO2 emissions can be associated with economic activities that lead to increased unemployment, raising the question of how to strike a balance between environmental sustainability and economic growth. Fig. 3 further separates countries by economic strength, indicating potential strategies for improving those that are disadvantaged.

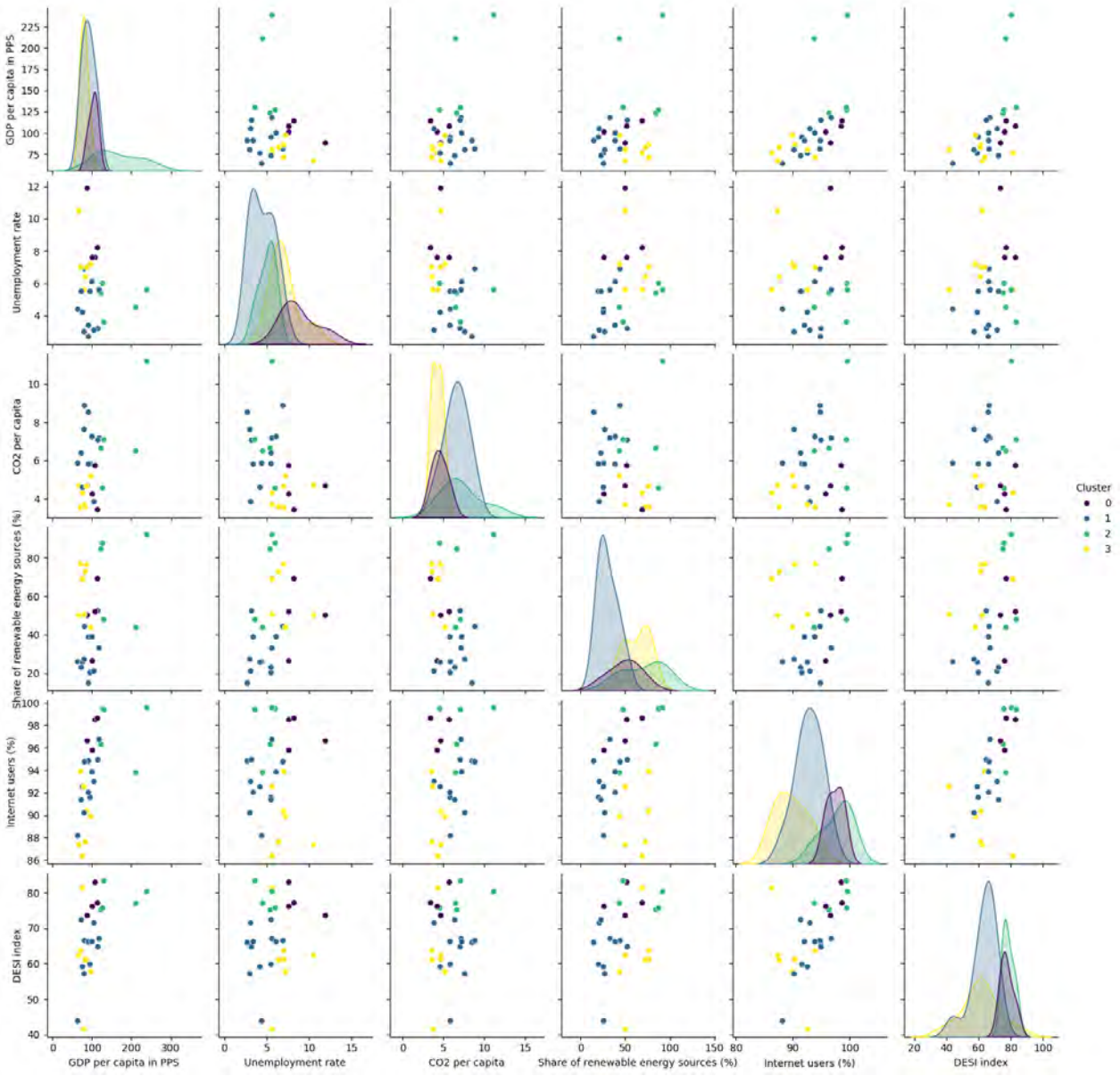


Figure 4 Pairplot of clusters

The graph in Fig. 4 shows a pairplot that visualizes the relationships between different economic

and social characteristics of countries, i.e. selected variables for the model, such as GDP per capita, unemployment rate, CO2 emissions, share of renewable energy sources, Internet users and DESI index, and the colors of the points indicate cluster groups.
The diagonal contains a histogram of the distribution of each feature, while the other graphs show the interrelationships between them, allowing the identification of similarities between countries within a cluster. This graph reveals several clusters of countries with clear connections between features, as well as differences in their characteristics.

It can be seen from the dispersion of clusters when observing GDP per capita in PPS and Internet users that a similar result of data clustering is repeated. When we have internet users on the ordinate and GDP per capita on the abscissa, we can see that the correlation is stronger, and vice versa when GDP per capita is on the ordinate and internet users on the abscissa, the points are very narrowly scattered.

Fig. 5 indicates that all observed countries have over 86% of internet users. It can be clearly seen that if the direction were drawn it would lean to the right. It directly shows that with the increase in GDP per capita, the number of internet users also increases. In addition to the two outliers shown in the second chart.

Clusters with greater economic power also have the highest percentage of Internet users. Then follows the next cluster with the average amount of Internet users and the average amount of GDP per capita, and the cluster with the lowest observed GDP per capita and the lowest percentage of Internet users.

In the case without outliers, the clusters are ideally arranged in relation to the centroids and the differences between the clusters are more clearly visible. If we look at the outliers, which fall into a separate cluster, they are actually outliers due to the GDP per capita variable.

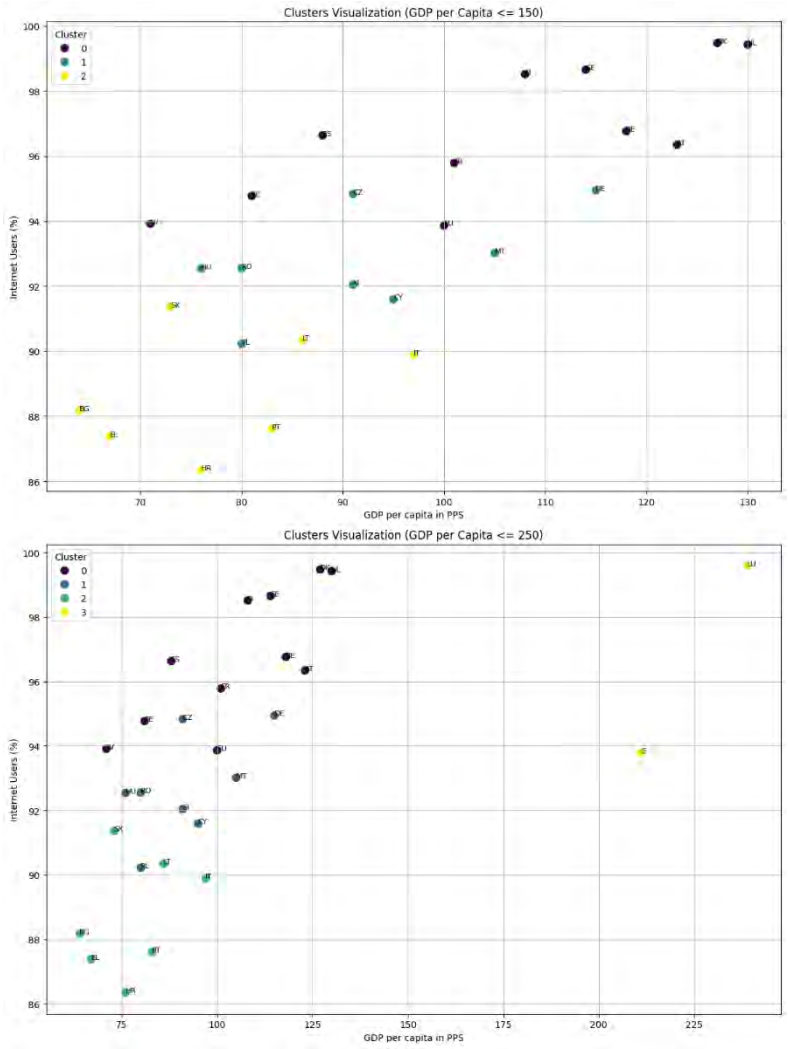


Figure 5 Clusters with reduced number of variables

The findings imply that higher GDP per capita correlates with greater internet usage, potentially influenced by factors like economic development and technological infrastructure. That also can be seen on Fig. 4 where the DESI index and GDP per capita make a graph. Countries with a higher DESI index have a higher GDP per capita because the cluster density increases with an increase in GDP per capita and the DESI index. Further exploration is necessary to understand each cluster's drivers behind these trends, for example which segments of technological infrastructure have been improved or invested in. That can significantly contribute to understanding how to make policies better and allocate investments to reach sustainable macroeconomic growth.

Conclusions

The analysis of the correlation matrix, as illustrated in Fig. 2, underscores the intricate relationships between digital transformation, economic performance, and environmental sustainability across EU countries. The findings reveal that PCA1 is significantly correlated with key economic indicators, most notably GDP per capita (0.86), the number of internet users (0.81), and the Digital Economy and Society Index (DESI) (0.77). This strong correlation indicates that countries excelling in digital transformation tend to have higher economic output and digital engagement among their populations, reinforcing the idea that digital tools can drive productivity and innovation.

Conversely, PCA2 correlates positively with the unemployment rate (0.84) and negatively with CO2 emissions per capita (-0.64). This suggests a complex interplay where regions with higher unemployment rates may also exhibit higher CO2 emissions, hinting at economic activities that are not aligned with sustainability. These correlations raise critical questions about the balance between economic growth and environmental responsibility, suggesting that policies aimed at reducing CO2 emissions might unintentionally affect employment in traditional sectors.

Moreover, the negative correlation between CO2 emissions and unemployment rate reinforces the need for strategic policy interventions. For instance, while reducing emissions is vital for sustainability, such efforts should be coupled with initiatives that create new job opportunities, particularly in green technologies and industries. This insight is crucial for formulating policies that not only aim for environmental goals but also prioritize economic stability and employment.

The findings from Fig. 2 highlight the necessity for EU policymakers to adopt a holistic approach in addressing the challenges of digital transformation and sustainable growth. By fostering a digital economy that supports environmental sustainability, EU countries can create a synergistic effect where economic performance, job creation, and ecological integrity are enhanced simultaneously.

In summary, the correlation analysis presents a compelling case for the integration of digital transformation strategies with macroeconomic and sustainability objectives. Correlations show the impact of digital transformation on the economy. In order to successfully integrate new technological achievements, it is necessary to analyze in more detail the successful countries in the application of digitization. Policymakers should leverage these insights to develop tailored interventions that address the unique characteristics of each member state, ultimately fostering a resilient and sustainable macroeconomic landscape across the European Union.

It should also be emphasized, as previously mentioned in the paper, that the strategy of reducing the carbon footprint can create new jobs that can offer greater added value to GDP.

References

[1] Prokop, C., & Tepe, M. (2022). Talk or type? The effect of digital interfaces on citizens' satisfaction with standardized public services. Public administration, 100(2), 427-443., https://doi.org/10.1111/padm.12739

[2] Kawane, T., Adu-Gyamfi, B., Cao, Y., Zhang, Y., Yamazawa, N., He, Z., & Shaw, R. (2024). Digitization as an Adaptation and Resilience Measure for MSMEs amid the COVID-19 Pandemic in Japan: Lessons from the Food Service Industry for Collaborative Future Engagements. Sustainability, 16(4), 1550., https://doi.org/10.3390/su16041550

[3] Sewpersadh, N. S. (2023). Disruptive business value models in the digital era. Journal of Innovation and Entrepreneurship, 12(1), 2., https://doi.org/10.1186/s13731-022-00252-1

[4] Criveanu, M. M. (2023). Investigating digital intensity and e-commerce as drivers for sustainability and economic growth in the EU countries. Electronics, 12(10), 2318., https://doi.org/10.3390/electronics12102318

[5] Škare, M., Gavurova, B., & Porada-Rochon, M. (2024). Digitalization and carbon footprint: Building a path to a sustainable economic growth. Technological Forecasting and Social Change, 199, 123045., https://doi.org/10.1016/j.techfore.2023.123045

[6] Asoy, E. (2024). A better proxy for technology determinant of economic growth: the international digital economy and society index. İstanbul İktisat Dergisi, 74(1), 225-251., https://doi.org/10.26650/ISTJECON2023-1327567

[7] Rollnik-Sadowska, E., & Bartkutė-Norkūnienė, V. (2024). Efektywność polityk rynku pracy w krajach UE. Comparative Economic Research, 27(2), 113., https://doi.org/10.18778/1508-2008.27.15

[8] Zeraibi, A., Radulescu, M., Khan, M. K., Hafeez, M., & Jahanger, A. (2024). Analyzing the linkage between public debt, renewable electricity output, and CO2 emissions in emerging economies: does the N-shaped environmental Kuznets curve exist?. Energy & Environment, 35(5), 2407-2430., https://doi.org/10.1177/0958305X231151678

[9] Oyewole, G. J., & Thopil, G. A. (2023). Data clustering: application and trends. Artificial Intelligence Review, 56(7), 6439-6475., https://doi.org/10.1007/s10462-022-10325-y

[10] Almais, A. T. W., Susilo, A., Naba, A., Sarosa, M., Crysdian, C., Tazi, I., ... & Wicaksono, H. (2023). Principal Component Analysis-Based Data Clustering for Labeling of Level Damage Sector in Post-Natural Disasters. IEEE Access, 11, 74590-74601., https://doi.org/10.1109/ACCESS.2023.3275852

Data centres and sustainability: the effect of geographical location and electricity mix through life cycle approach perspective

Jelena Topić Božič[1,2], Simon Muhič[2,3,1]
[1] Rudolfovo – Science and Technology Centre Novo mesto
Podbreznik 15, 8000 Novo mesto, Slovenia
[2] Faculty of Industrial Engineering Novo mesto
Šegova ulica 112, 8000 Novo mesto, Slovenia
[3] Institute for Renewable Energy and Efficient Exergy Use, INOVEKS d.o.o,
Cesta 2. grupe odredov 17, 1295 Ivančna Gorica, Slovenia

jelena.topic.bozic@rudolfovo.eu,simon.muhic@fini-unm.si

Abstract: *Operating computers, networks, and data centers involve use large amounts of electricity. Data centers and data transmission networks each account for 1-1.5% of global electricity use. Their environmental impact will depend on the geographical location and the electricity mix. This study aimed to illustrate the effect of geographical location on the environmental impact of electricity use of data center to showcase how a difference in electricity mix profile can impact the GHG emissions of a data center. The data for Slovenia and Serbia demonstrates the effect of vastly different electricity mixes, as these two countries' electricity profiles vary greatly. In the case of the electricity use of 50 MWh, the life cycle electricity GHG emission factor is 4.4 times higher in Serbia compared to Slovenia due to the main share of thermal power plants.*

Key Words*: data centers, sustainability, GHG emissions, life cycle approach, electricity use*

1 Introduction

Digital information has encountered explosive growth, and data centers present the core infrastructure and the information backbone of an increasingly digitalized world [1–3]. The Information and Communication Technology (ICT) sector has gained much attention in connection with climate change; as our dependency on digital services grows, so do the concerns that the ICT sector's GHG emissions may follow [4]. Digital services such as video streaming, social media, big data, cryptocurrency mining, artificial intelligence, and digitalization of business processes and production flows are leading to more and more data being stored and processed in data centers, resulting in the dynamic growth of data center capacities [5, 6]. Data centers are carbon-intensive enterprises due to their massive energy use, and it is estimated that the data center industry will account for 8% of global carbon emissions by 2030 [7]. Malmodin in Lundén [8] estimated that the IT sector's carbon footprint is 730 Mt CO_2-equivalents (CO_2-eq) or 1.4% of global emissions. Recently, Freitag et al. [9] stipulated that the proportion of the IT sector's share of global greenhouse gas (GHG) emissions could be around 2.1–3.9% when the full life-

cycle and supply chain of IT products and infrastructure are considered. Despite the extensive technological development we encounter every day, the life cycle of IT equipment is associated with several negative environmental impacts. The generation of e-waste is expected to increase to 74.7 Mt in 2030. In addition, the operation of computers, networks, and data centers involves use of large amounts of electricity [6]. Data centers and data transmission networks each account for 1-1.5% of global electricity use [10].

The concept of sustainable information technologies has gained increasing attention in recent years due to the growing awareness regarding sustainability and the environment. According to the Green Software Foundation (GSF) the first principle goal is to emit as little carbon (GHG) as possible per unit of work. Therefore, developing green software means producing software that uses the least energy. Measuring an application's energy use and evaluating the efficiency of hardware use is essential. The energy used has different impacts based on carbon intensity. The latter varies depending on where and when the relative energy is used [6]. The energy efficiency of data centers is constantly increasing, to the point that an increase in computing workload of 550% from 2010 to 2018 has resulted only in a rise in energy demand of 6% [11, 12].

As data centers use a lot of electricity, their environmental impact will depend on the geographical location and electricity mix. The GHG emissions factors for electricity use can be determined on a national level following an activity-based (IPCC) approach and life cycle approach) [13]:

- Activity based approach (IPCC):
    - CO_2 emission factor is provided only for CO_2 emissions (in tCO_2/MWh)
    - GHG emissions factor is provided for CO_2, CH_4 and N_2O (in tCO_2-eq/MWh)
- Life cycle approach (LC):
    - GHG emissions factor is provided, accounting for GHG emissions, including CO_2, N_2O and CH_4 (in tCO_2-eq/MWh), including upstream (supply chain) emissions.

This study aimed to illustrate the effect of geographical location on the environmental impact of electricity use of data center to showcase how a difference in electricity mix profile can impact the GHG emissions of a data center.

2 Methodology

The dataset for GHG emissions factors for national electricity from the EU Joint Research Centre was used. The dataset is available online for public use [14]. Three types of emission factors can be found in this dataset, following two approaches:

- An activity-based (IPCC) approach and a life-cycle (LC) approach for EU countries, Clima-MED countries, Western Balkan & Türkiye, Other European countries (Switzerland, UK), and Other countries (i.e. Kazakhstan);
- In the activity-based approach, an emission factor is provided for only CO_2 emissions (in tonnes CO_2/MWh), and another emission factor is provided for GHG emissions, namely CO_2, CH_4, and N_2O (in tonnes CO_2-eq/MWh);
- in the LC approach, an emission factor is provided accounting for GHG emissions, namely CO_2, N_2O, and CH_4 (in tonnes CO_2-eq/MWh), including upstream (supply chain) emissions [14].

The data for Slovenia and Serbia demonstrate the effect of vastly different electricity mixes as the electricity profile of these two countries varies greatly. An estimation of 50 MWh of electricity use for a data center was used to calculate emission factors.

Table 1: Emission factors for national electricity use for 2014 - 2021 *[14]*.

Approach	Unit/Year	2014	2015	2016	2017	2018	2019	2020	2021
Slovenia									
IPPC CO_2	tCO_2/MWh	0.227	0.250	0.252	0.234	0.218	0.218	0.213	0.203
IPCC GHG	tCO_2-eq/MWh	0.228	0.251	0.253	0.236	0.219	0.219	0.214	0.205
LC GHG	tCO_2-eq/MWh	0.257	0.286	0.287	0.270	0.249	0.250	0.242	0.235
Serbia									
IPPC CO_2	tCO_2/MWh	0.879	1.046	1.026	1.023	0.963	0.972	1.017	/
IPCC GHG	tCO_2-eq/MWh	0.881	1.049	1.029	1.026	0.966	0.974	1.019	/
LC GHG	tCO_2-eq/MWh	0.903	1.074	1.054	1.052	0.991	1.001	1.046	/

3 Results

3.1 The electricity production profile of Slovenia and Serbia

In Fig1, electricity generation source profiles for Slovenia and Serbia are shown, which vary greatly between countries. The electricity in Serbia is mainly produced from thermal power plants (63.2%), followed by hydropower plants (31.3%), with wind power having 2.8% share.

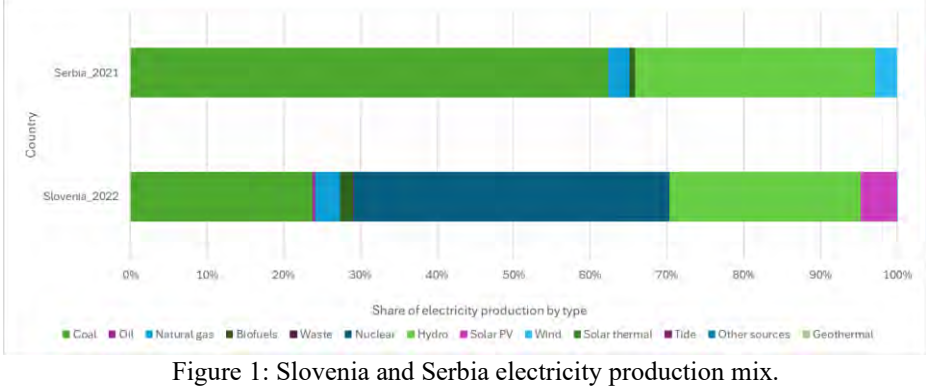


Figure 1: Slovenia and Serbia electricity production mix.

Slovenia's thermal power plant electricity production accounts for 23.7% share, while nuclear (41.1%) and hydropower (25.0%) are the two sources with the highest share. Additionally, solar power in Slovenia has 4.7% share [15, 16].

3.2 GHG emission factors for electricity use in Slovenia and Serbia

Fig 2 shows the emission factors from electricity use. It can be seen that LC GHG approach yields higher values in both countries as it also includes the supply chain emissions. The GHG emissions from electricity generation depend on multiple factors, including conversion technology type, plant efficiency, fuel type and quality, fuel transport distance, local conditions (such as wind and solar resources) etc. [17]. Compared to Slovenia, Serbia has significantly higher emissions from electricity due to the main share of coal in the electricity mix profile. In 2020, the LC GHG emissions factor was 4.4 times lower for MWh of electricity produced in Slovenia than Serbia. Slovenia has seen a slow decline in emission factors in 2014 – 2021 due to a larger share of renewable energy sources.

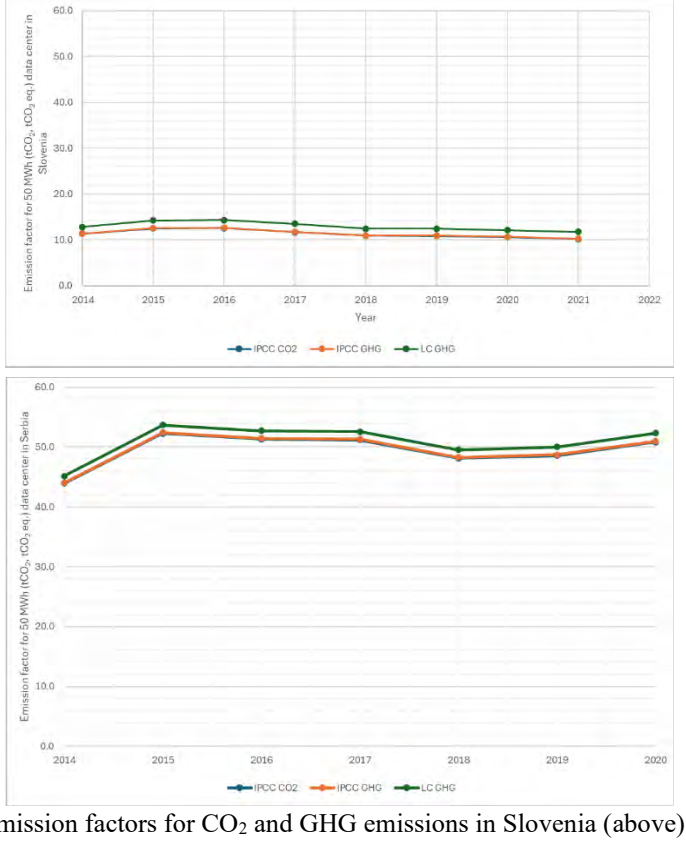


Figure 2: Emission factors for CO_2 and GHG emissions in Slovenia (above) and Serbia (below) for a data center using 50 MWh of electricity.

The growth of data center services and the increase in electricity use highlight the need to pursue energy efficiency opportunities [18]. The carbon footprint of a data center is directly proportional to its efficiency and the carbon intensity of the location, which is the most crucial actor in terms of the total carbon footprint, which varies from less than 20 gCO_2-eq/kWh in Norway and Switzerland to over 800 gCO2-eq/kWh in Australia, South Africa and some parts of the US [19]. Based on the data from Rong et al. [1], servers and cooling systems present the most substantial energy-draining facilities in data centers. They account for a dominant share of the total operating costs. Reducing energy use for

servers and cooling systems is the key issue of the sustainable development of data centers. The energy saving measures will contribute differently to the energy saving to varying stages of data centers globally. The location (i.e., site selection of data centers) may lead to almost 12-15% off in total energy use. The construction of data centers may reduce energy use by 25-30%, especially when choosing low-power servers and auxiliary energy-saving devices. The operators of data centers can save about 10-15% of the total energy use by optimizing resource scheduling algorithms and management strategies [1]. The International Energy Agency (IEA) stipulates that more efforts are needed to track data centers and data transmission networks and the main recommendations are [10]:

- *Improvement of data collection and transparency*: sharing information about energy use can help inform energy analysis and policy making. National research programs can develop better modeling tools to improve understanding and forecasting of data centers and networks' energy and sustainability impacts.
- *Enactment of policies that encourage energy efficiency, demand response, and clean energy procurement*: Governments can offer guidance, incentives, and standards to promote further energy efficiency, while regulations and price signals could help incentivize demand-side flexibility.
- *Support of the utilization of waste heat from data centers*: Waste heat from data centers could help heat nearby commercial and residential buildings or supply industrial heat users, reducing energy use from other sources.
- *Reduction of life cycle environmental impacts*: data centers and data transmission networks are also responsible for "embodied" life cycle emissions from raw material extraction, manufacturing, transport, and end-of-life disposal or recycling.

According to Zhu et al., there is a need for highly efficient energy-conservation and carbon-emission technologies and strategies for effective control of data centers' cooling systems [20]. The main decarbonization strategies identified are [20]:

- Optimization of IT equipment,
- Research and development of advanced cooling technologies;
- The single energy-saving method is difficult to meet the goal of zero-carbon emission due to differences in geographical location, natural resources and economic basis.
-

Strategies to achieve sustainability in data centers have been through the use of renewable energy sources and waste heat recovery. From the software point of view, virtualization and optimal scheduling of algorithms is seen as a strategy to reduce energy use in data centers [21].

4 Conclusions

Currently, most efforts regarding the sustainability of data centers are focused on energy efficiency. However, looking toward the future, the focus on sustainable data center operations will shift from the sole consideration of energy requirements to other sustainability indicators. Even though the main focus has been on GHG emissions, other indicators such as water scarcity should also be considered following the life cycle approach, which gives information through the whole data center life cycle. Because of the increasing energy use, the topic of waste heat utilization will play a central role and should be considered when selecting locations for data centers. Two main approaches to improve the sustainability of data centers are increasing energy efficiency and re-using

the energy wasted. The application of renewable sources and load scheduling should be considered to achieve sustainability goals. Strategically locating new data centers can reduce their environmental footprint as climatic factors can make some areas more favorable due to lower ambient temperatures, thus reducing cooling requirements which leads to reduction in both direct and indirect water consumption, as well as GHG emissions, associated with data center operation. Differences between countries are significant in terms of electricity mixes and resulting GHG emissions, with Serbia having five times higher emission factor per 1 kWh of electricity compared to Slovenia. As most data centers meet their electricity demands from the grid, the electricity mix plays a significant role in a data center's environmental footprint.

5 References

[1] RONG, Huigui, Haomin ZHANG, Sheng XIAO, Canbing LI a Chunhua HU. Optimizing energy consumption for data centers. *Renewable and Sustainable Energy Reviews* [online]. 2016, **58**, 674–691. ISSN 1364-0321. Available at: doi:10.1016/j.rser.2015.12.283

[2] MASANET, Eric, Arman SHEHABI, Nuoa LEI, Sarah SMITH a Jonathan KOOMEY. Recalibrating global data center energy-use estimates. *Science* [online]. 2020, **367**(6481), 984–986. Available at: doi:10.1126/science.aba3758

[3] DEPOORTER, Victor, Eduard ORÓ a Jaume SALOM. The location as an energy efficiency and renewable energy supply measure for data centres in Europe. *Applied Energy* [online]. 2015, **140**, 338–349. ISSN 0306-2619. Available at: doi:10.1016/j.apenergy.2014.11.067

[4] MALMODIN, Jens, Nina LÖVEHAGEN, Pernilla BERGMARK a Dag LUNDÉN. ICT sector electricity consumption and greenhouse gas emissions – 2020 outcome. *Telecommunications Policy* [online]. 2024, **48**(3), 102701. ISSN 0308-5961. Available at: doi:10.1016/j.telpol.2023.102701

[5] HINTEMANN, Ralph a Simon HINTERHOLZER. *Energy consumption of data centers worldwide How will the Internet become green?* 2020.

[6] PAZIENZA, Andrea, Giovanni BASELLI, Daniele Carlo VINCI a Maria Vittoria TRUSSONI. A holistic approach to environmentally sustainable computing. *Innovations in Systems and Software Engineering* [online]. 2024, **20**(3), 347–371. ISSN 1614-5054. Available at: doi:10.1007/s11334-023-00548-9

[7] CAO, Zhiwei, Xin ZHOU, Han HU, Zhi WANG a Yonggang WEN. Toward a Systematic Survey for Carbon Neutral Data Centers. *IEEE Communications Surveys & Tutorials* [online]. 2022, **24**(2), 895–936. ISSN 1553-877X. Available at: doi:10.1109/COMST.2022.3161275

[8] MALMODIN, Jens a Dag LUNDÉN. The Energy and Carbon Footprint of the Global ICT and E&M Sectors 2010–2015. *Sustainability* [online]. 2018, **10**(9), 3027. ISSN 2071-1050. Available at: doi:10.3390/su10093027

[9] FREITAG, Charlotte, Mike BERNERS-LEE, Kelly WIDDICKS, Bran KNOWLES, Gordon S. BLAIR a Adrian FRIDAY. The real climate and transformative impact of ICT: A critique of estimates, trends, and regulations. *Patterns* [online]. 2021, **2**(9), 100340. ISSN 2666-3899. Available at: doi:10.1016/j.patter.2021.100340

[10] IEA. Data centres & networks. *IEA* [online]. 2024 [vid. 2024-09-30]. Available at: https://www.iea.org/energy-system/buildings/data-centres-and-data-transmission-networks

[11] SIDDIK, Md Abu Bakar, Arman SHEHABI a Landon MARSTON. The

environmental footprint of data centers in the United States. *Environmental Research Letters* [online]. 2021, **16**(6), 064017. ISSN 1748-9326. Available at: doi:10.1088/1748-9326/abfba1

[12] FARFAN, Javier a Alena LOHRMANN. Gone with the clouds: Estimating the electricity and water footprint of digital data services in Europe. *Energy Conversion and Management* [online]. 2023, **290**, 117225. ISSN 0196-8904. Available at: doi:10.1016/j.enconman.2023.117225

[13] BASTOS, Joana, Matteo Giacomo PRINA a Rita GARCIA. Life-cycle assessment of current and future electricity supply addressing average and marginal hourly demand: An application to Italy. *Journal of Cleaner Production* [online]. 2023, **399**, 136563. ISSN 0959-6526. Available at: doi:10.1016/j.jclepro.2023.136563

[14] BASTOS, Joana, Fabio MONFORTI-FERRARIO a Giulia MELICA. GHG Emission Factors for Electricity Consumption. *JRC Publications Repository* [online]. 8. únor 2024 [vid. 2024-09-18]. Available at: https://publications.jrc.ec.europa.eu/repository/handle/JRC136340

[15] INTERNATIONAL ENERGY AGENCY (IEA). Energy system of Slovenia. *IEA* [online]. 2024 [vid. 2024-04-15]. Available at: https://www.iea.org/countries/slovenia

[16] INTERNATIONAL ENERGY AGENCY (IEA), IEA. Energy system of Serbia. *IEA* [online]. 2024 [vid. 2024-04-15]. Available at: https://www.iea.org/countries/serbia

[17] SCARLAT, Nicolae, Matteo PRUSSI a Monica PADELLA. Quantification of the carbon intensity of electricity produced and used in Europe. *Applied Energy* [online]. 2022, **305**, 117901. ISSN 0306-2619. Available at: doi:10.1016/j.apenergy.2021.117901

[18] SHEHABI, Arman, Eric MASANET, Hillary PRICE, Arpad HORVATH a William W. NAZAROFF. Data center design and location: Consequences for electricity use and greenhouse-gas emissions. *Building and Environment* [online]. 2011, **46**(5), 990–998. ISSN 0360-1323. Available at: doi:10.1016/j.buildenv.2010.10.023

[19] BOLÓN-CANEDO, Verónica, Laura MORÁN-FERNÁNDEZ, Brais CANCELA a Amparo ALONSO-BETANZOS. A review of green artificial intelligence: Towards a more sustainable future. *Neurocomputing* [online]. 2024, **599**, 128096. ISSN 0925-2312. Available at: doi:10.1016/j.neucom.2024.128096

[20] ZHU, Hongyu, Dongdong ZHANG, Hui Hwang GOH, Shuyao WANG, Tanveer AHMAD, Daijiafan MAO, Tianhao LIU, Haisen ZHAO a Thomas WU. Future data center energy-conservation and emission-reduction technologies in the context of smart and low-carbon city construction. *Sustainable Cities and Society* [online]. 2023, **89**, 104322. ISSN 2210-6707. Available at: doi:10.1016/j.scs.2022.104322

[21] AHWAZI, Amin Ziagham, Chiara BORDIN, Sambeet MISHRA, Alexander HORSCH a Phuong Hoai HA. Sustainable and decarbonized data-center facilities: A socio-techno-economic discussion. In: *2021 IEEE PES Innovative Smart Grid Technologies - Asia (ISGT Asia)*: *2021 IEEE PES Innovative Smart Grid Technologies - Asia (ISGT Asia)* [online]. 2021, s. 1–5 [vid. 2024-10-28]. ISSN 2378-8542. Available at: doi:10.1109/ISGTAsia49270.2021.9715647

Emerging trends in generative artificial intelligence: Insights from patent analysis using Lens.org toolkit

Mirjana Pejić Bach, Jasmina Pivar
Unitversity of Zagreb, Faculty of Economic and Business
Trg J. F. Kennedyja 6, 10000 Zagreb, Croatia
{mpejic, jpivar}@net.efzg.hr

Robert Fabac
Unitversity of Zagreb, Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Croatia
{rfabac}@foi.hr

Abstract: *Generative artificial intelligence (GenAI) has leveraged rapid change across all domains of human society. This paper investigates the emerging trends in GenAI using the patent analysis approach, which is the process of analysing patent documents to uncover innovation trends in a particular technology field. We utilise the Lens.org toolkit that merges patents and scholarly work on an open web platform. Results indicate a boom of patenting in GenAI in 2023 and 2024, with most patents in a legal status 'pending'. The top applicants are Microsoft and Adobe, and most of the applicants are from China and USA. The most frequently protected technical content are generative networks and learning methods. Association rule analysis was used to detect content that is often protected together, such as query formulation, querying, and result presentation. Additionally, it reveals how advancements in areas like probabilistic network architectures and convolutional networks are driving progress in adversarial learning.*

Key Words*: generative artificial intelligence, Gen AI, patent analysis, Lens.org, text mining, patent landscape*

1 Introduction

The development of Generative Artificial Intelligence (Gen AI) has evolved significantly over the past decade, with key milestones shaping its trajectory. One pivotal step was the advent of deep learning in the early 2010s, characterised by the use of convolutional neural networks (CNNs). Notable models like AlexNet demonstrated impressive performance in image classification, paving the way for more advanced generative models [5]. The introduction of Generative Adversarial Networks (GANs) by Goodfellow et al. [1] marked a breakthrough, enabling the generation of high-quality synthetic images. Additionally, an important category of generative models, Variational Autoencoders (VAEs), emerged during this period [4]. The next phase was ushered in by the transformer architecture [11], which facilitated the development of large-scale language models. During these years, models like GPT [9] demonstrated the ability to understand and generate human-like text. These models, leveraging massive datasets, achieved remarkable performance across various generative tasks. Today, Gen AI encompasses a wide range of applications, extending beyond text generation to include image synthesis, music composition, and more.

Currently, GANs technology plays a pivotal role, particularly in tasks related to image and video generation. GANs have revolutionised fields like healthcare, art, and

entertainment through innovations in image synthesis, style transfer, and data augmentation. Key challenges of GANs include training instability, where models often face issues like non-convergence and mode collapse, leading to limited output diversity [1], [10]. Additionally, adversarial examples generated by GANs can mislead models, raising concerns about robustness and security [15]. GANs models were recently challenged by Denoising Diffusion Probabilistic Models introduced by Ho et al. [3]. Recent research demonstrated that diffusion-based methods could outperform GANs on image generation tasks, particularly in stability and diversity [8], [15].

Since GenAI is a subset of AI that specifically focuses on algorithms and models for generating new content— the patent areas for AI and GenAI differ significantly. In the main category, AI patents focus on algorithms such as machine learning, computer vision, and natural language processing (NLP), emphasising improvements in efficiency, scalability, and accuracy. In contrast, GenAI patents target models that generate new content, including GANs and large language models (LLMs), often involving novel architectures or training techniques for producing high-quality synthetic data, text, and images. AI patents have a broad range of applications across industries such as healthcare (for diagnostics), predictive analytics, and financial trading. Computer vision-related patents include tasks like facial recognition, autonomous driving (e.g., object detection), and security systems (e.g., surveillance). On the other hand, GenAI patents focus on content generation, encompassing text generation (e.g., chatbots and writing), image and video synthesis, and audio generation (e.g., music composition and speech synthesis).

The use of patent analysis to explore emerging technological trends can provide valuable insights into the direction of AI innovations and the domains of their applications. Key expectations from such analysis include identifying leading innovators, tracking geographical trends, predicting technological roadmaps, and understanding commercialisation potential. Gen AI patent documents (GenAI PDs) started in 2003, with a threshold of 200 patent documents in 2024. McKinsey's study in 2023 identifies numerous applications of Gen AI across 16 business functions, which could potentially generate up to $4.4 trillion in annual economic benefits when implemented across various industries [7]. This figure, in addition, represents a 15% to 40% increase over the estimated economic value that non-generative AI and analytics can unlock. Additionally, Gen AI has been shown to perform over 2,100 detailed work activities across more than 800 occupations. Overall, this highlights an economic potential of $17.1 trillion to $25.6 trillion for AI contributions [7].

This research aims to gain comprehensive insights into the current state of Gen AI patents globally, with a particular focus on the technical content found within patent documents. By identifying key emerging trends related to patent families, companies, and geographical distributions, it becomes possible to forecast future applications of Gen AI, track the development of specific technologies, and assess the impact of Gen AI models on various industries in different countries.

In the following parts of the paper, we first provide an overview and explanation of the applied methodology. Then, in the results section, we determine the characteristic timeline of Gen AI patent applications, identify the top applicants, extract key themes, and analyse the technical content protected in these GenAI PDs. This is followed by a discussion of the results and a chapter on conclusions.

2 Methodology

First, the patent selection was made using the Lens toolkit, a free platform that enables patent searching and analysis from global databases (Lens.org, 2024). The following search string was used to search the Lens database on 30 September 2024: Title: ("generative artificial intelligence") OR (Title: ("generative AI") OR (Abstract: ("generative artificial intelligence") OR Abstract: ("generative AI"))). In that manner, we tracked the patent documents that mentioned the terms “generative AI” or “generative artificial intelligence” in their titles or abstracts. Next, we filtered the results to include only simple patent families, which originate from the initial patent document known as the priority documents. A simple patent family is a collection of patent documents that safeguard individual innovations across multiple countries. Therefore, the number of extracted simple patent families reflects the actual number of distinct patents in a field without duplicating the same patent documents in various countries. The authors extracted 271 simple patent families. We detected four of those simple patent families being duplicates. Therefore, they were excluded from the analysis, which left us with a total of 267 simple patent families, which were the basis for further analysis.

Second, analysing patents followed, often referred to as the "patent landscape." First, the timeline is analysed using data on publication year, which indicates the date on which the respective patent office publishes the patent. Second, the authors analysed the main applicants for patents in the field. Third, patents were distributed according to jurisdictions, and the type of patent documents were described.

Third, a technical content analysis of GenAI PDs was conducted, which involved analysing the titles of the patents using a text-mining approach and conducting an in-depth analysis of the patents according to the International Patent Classification (IPC). To explore the content of GenAI PDs, authors have extracted frequent phrases and topics from GenAI PDs. The extraction was made based on GenAI PDs’ titles, abstracts and keywords, using WordStat Provalis ver 8. Stemming, lemmatisation, and exclusion of words with little semantic value were conducted as part of text preprocessing before performing the extraction of phrases and topics. Finally, the technical content of GenAI PDs was analysed according to the International Patent Classification (IPC) system, which assigns IPC codes to patent documents to indicate the protected technical content. In this research, the analysis of the patent documents has been conducted according to the group level of the IPC system, which is the most detailed level of the IPC system. Finally, we used association rules to conduct and analyse the IPC codes, with the goal of identifying the heterogeneity of technical content in patents related to Gen AI.

3 Results

3.1 Timeline of Gen AI patent documents

The first simple patent families with the term “generative artificial intelligence” or “generative AI” in title, abstract or claim were protected in 2003 and 2004 with the title “Methods and systems for automated property valuation”, which are both currently discontinued. Although both had the same title, they were different in their object of protection. The first patent, applied in 2003, involved training a fuzzy-neural network to estimate property values with a system based on fuzzy logic and neural network rules, while the second patent, applied in 2004, combined multiple valuation processes, such as AI-driven fuzzy-neural networks for property value assessment. After that, for 18 years,

there were no patents in the domain of Gen AI, and the first patent after this long break was the patent "Content generation using target content derived modelling and unsupervised language modelling", legally protected in 2021. Table 1 presents the current legal status of GenAI PDs, among which most are still pending (68.91%) and were mostly applied in 2024. This indicates a striking trend of rapid patenting in the Gen AI domain. Active patents account for 27.34%, while a relatively low number of patents are discontinued, unknown, or expired. This distribution highlights that the majority of GenAI patent applications have been filed in the last two years, reflecting the rapid development and interest in protecting intellectual property in this emerging field. In 2024, the boom of patenting applications in the Gen AI domain emerged, with a total of 190 patent families, among which 134 are pending, 50 are active, and the rest are unknown or discontinued.

Table 1: Gen AI patent documents according to legal status – until 30 September 2024

Legal status	Year				# of Simple Patent Families	Structure in %
	<=2021	2022	2023	30 Sept., 2024		
pending	0	1	49	134	184	68,91%
active	1	2	20	50	73	27,34%
unknown	0	0	0	5	5	1,87%
discontinued	0	0	2	1	3	1,12%
expired	2*	0	0	0	2	0,75%
Total	3	3	71	190	267	100,00%

Note: * two patents expired in 2003 and 2004

3.2 Applicants of Gen AI patent documents

Table 2 shows key applicants who are working in the technological field of Gen AI. Microsoft and Adobe are leaders in the field. Some of the Gen PDs applied by Microsoft are related to generative networks, such as patents titled "*Artificial intelligence prompt processing and storage system"* or "*Generative artificial intelligence development system".* On the other hand, Adobe applied several GenAI PDs related to image data processing or generation in general, such as patents titled "*Patterned glyph using Gen AI"* or "*Target scene synthesis using generative AI".*

Table 2: Top 10 Applicants of GenAI PDs – *until 30 September 2024*

Top 10 Applicants	# of Simple Patent Families
MICROSOFT TECH LICENSING LLC	11
ADOBE INC	8
SAV COM LLC	6
URP CO LTD	6
ALIPAY HANGZHOU INF TECH CO LTD	6
MCCARSON BRIAN	4
THE MATRIX CO LTD	4
BEIJING BAIDU NETCOM SCI & TECH CO LTD	4
JINAN MINGQUAN DIGITAL COMMERCE CO LTD	4
ZHEJIANG ZEEKR INTELLIGENT TECH CO LTD; GEELY HOLDING GROUP CO LTD	4

Table 3 presents the publication timeline of Gen AI patents across jurisdictions. In the current year, 2024, among the observed jurisdictions, most GenAI PDs have been published in China and the fewest in Taiwan. These results indicate a recent surge in Gen AI.

Table 3: GenAI PDs according to jurisdiction over time – *until 30 September 2024*; Note: CN - China, US - the United States, KR - South Korea, WO - the World Intellectual Property Organization, GB - Great Britain, AU - Australia, EP - the European Patent Organization, CA - Canada, and TW - Taiwan

Jurisdiction / Year	CN	US	KR	WO	GB	AU	EP	CA	TW	Total
>=2021	0	1	0	0	0	0	0	0	0	1
2022	1	1	0	1	0	0	0	0	0	3
2023	48	12	5	2	2	0	1	0	1	71
30 September, 2024	70	73	25	13	5	2	1	1	0	190
Total	119	87	30	16	7	2	2	1	1	265

3.3. Topics of Gen AI patent documents

Table 4 shows the most frequent phrases in Gen AI PDs' titles extracted by Wordstat that occurred in 20 or more GenAI PDs. The keywords that were used in the search string ("generative artificial intelligence" and "generative AI") were excluded from the phrase list to derive more focused results. The most frequent phrases are systems and methods, data set, natural language and real-time. Most of the phrases indicate the technology being used, such as natural language, deep learning, machine learning, and neural networks. At the same time, only a few were related to the application domain, such as operation and maintenance, electronic equipment, and technical data. This rather generic result is expected since we focused on the most frequent phrases.

Table 4: Phrases in GenAI PDs (20+ frequency)

Phrase	FREQ	NO. CASES	% CASES	TF • IDF
SYSTEMS AND METHODS	53	35	13,11%	46,8
DATA SET	42	13	4,87%	55,1
NATURAL LANGUAGE	41	23	8,61%	43,7
REAL TIME	37	21	7,87%	40,9
REAL ESTATE	28	1	0,37%	67,9
DEEP LEARNING	27	13	4,87%	35,4
MACHINE LEARNING	23	16	5,99%	28,1
NEURAL NETWORK	22	13	4,87%	28,9
AI MODEL	21	10	3,75%	30
MULTI MODAL	21	12	4,49%	28,3
OPERATION AND MAINTENANCE	21	4	1,50%	38,3
COMPUTING SYSTEM	20	8	3,00%	30,5
ELECTRONIC EQUIPMENT	20	16	5,99%	24,4

To gain more in-depth insight into the topic of GenAI PDs, we have used WordStat ver 8 for topic extraction. Table 5 identifies key topics extracted from GenAI PDs, revealing a strong focus on "Real-Time" operations and "Machine Learning" interfaces as central themes in AI innovation. The topic "Response Generation" has the highest frequency, suggesting a major focus on generating AI-driven responses to user prompts and queries. Other significant topics include systems and methods for automation, natural language processing, and AI model training, emphasising the technological advancements in real-time AI applications, machine learning, and model development.

Table 5: Topics extracted from Gen AI patent documents

#	TOPIC	KEYWORDS	FREQ	CASES
1	REAL	REAL; TIME; MAINTENANCE; OPERATION; EFFICIENCY; IMPROVED; ACCURACY; CONTROL; INVENTION; TECHNOLOGY; FIELD; LARGE; METHOD COMPRISES; TECHNICAL FIELD; DATA SET; FIELD OF ARTIFICIAL INTELLIGENCE; LARGE LANGUAGE MODEL; ELECTRONIC EQUIPMENT AND A STORAGE; REAL TIME; BASED ON A GENERATIVE ARTIFICIAL; DEEP LEARNING;	224	89
2	MACHINE LEARNING INTERFACE	MACHINE; INTERFACE; LEARNING; CONFIGURED; PROCESS; DIGITAL; COMMUNICATION; PLURALITY; MODELS; BASED; MACHINE LEARNING; DEEP LEARNING; USER INTERFACE; LEARNING MODEL; AI MODEL;	206	108
3	QUESTION AND ANSWER	ANSWER; QUESTION; COMPUTER; KNOWLEDGE; FIELD; ANALYSIS; RESULT; OBTAINING; TECHNICAL FIELD; PREDICTION RESULT; QUESTION AND ANSWER; ANALYSIS METHOD; COMPUTER PROGRAM;	108	46
4	SYSTEMS AND METHODS	METHODS; SYSTEMS; AUTOMATICALLY; RELATED; GENERATING; SYSTEMS AND METHODS; SYSTEMS AND METHODS FOR AUTOMATICALLY; EXPERTISE THAT IS TRADITIONALLY NEEDED; GENERATING PREDICTIVE; GENERATIVE AI SYSTEMS AND METHODS; GENERATIVE ARTIFICIAL INTELLIGENCE THAT REQUIRES; REQUIRE THE TECHNICAL SKILLS; GENERATIVE ARTIFICIAL INTELLIGENCE PIPELINES; METHODS AND SYSTEMS; SUGGESTING A RELATED;	152	70
5	NATURAL LANGUAGE	DATABASE; NATURAL; NATURAL LANGUAGE;	101	8
6	MODULE	MODULE; MANAGEMENT; INTELLIGENT; VEHICLE; SYSTEM; CONTROL; SYSTEM BASED; SYSTEM COMPRISES; MANAGEMENT SYSTEM; METHOD AND SYSTEM BASED; COMPUTING SYSTEM; VEHICLE INFORMATION; GENERATIVE AI MODULE; UTILIZING GENERATIVE ARTIFICIAL INTELLIGENCE;	207	68
7	RESPONSE GENERATE	RESPONSE; GENERATE; LANGUAGE; PROMPT; ASSOCIATED; USER; QUERY; NATURAL; CONTENT; RECEIVING; MODEL; GENERATED; BASED; TEMPLATE; GENERATING; INPUT; SEARCH; AI MODEL;	697	212
8	IMPROVED	IMPROVED; EQUIPMENT; OBTAINING; ACCORDING; PRESET; TARGET; OBTAIN; INVENTION; INITIAL; ELECTRONIC; REQUEST; EFFICIENCY; METHOD COMPRISES; TECHNICAL FIELD; ELECTRONIC EQUIPMENT; DATA SET; ELECTRONIC EQUIPMENT AND A STORAGE; FIELD OF ARTIFICIAL INTELLIGENCE; MODEL TO OBTAIN; ACCORDING TO THE INVENTION; INFORMATION TO OBTAIN;	329	119
9	UNIT	UNIT; SERVER; PROVIDING; GENERATES; SERVICE; DISCLOSED; TERMINAL; GENERATION UNIT; METHOD FURTHER INCLUDES GENERATING; SERVER INCLUDES;	158	63

10	TRAINING	SERVICE USING GENERATIVE; USER TERMINAL; PRESENT DISCLOSURE; AI MODEL; ACCORDING TO AN EMBODIMENT; TRAINING; DATA; FEATURE; PERFORMING; NETWORK; ACCURACY; SET; OBTAIN; MODEL; HIGH; TRAINED; OBTAINED; OBTAINING; DEEP; INPUTTING; RESULT; GENERATIVE; TEXT; GENERATIVE ARTIFICIAL; AI MODEL;	650	171

3.4. Technical content protected together with Gen AI patent documents

Figure 1 shows the most frequently protected content by GenAI PDs according to the IPC system. The most frequently protected content of GenAI PDs is computing arrangements based on biological models, particularly Generative networks (G06N3/0475, 48 simple patent families). Generative networks - Learning methods (G06N3/08) were protected by 38 simple patent families. A significant number of patent families were related to machine learning (G06N20/00; 35 simple patent families). Query formulation (G06F16/332, 31 simple patent families) and querying (G06F16/33, 20 simple patent families) as part of file system structures were registered as well. Handling natural language data with semantic analysis (G06F40/30) was part of 24 simple patent families.

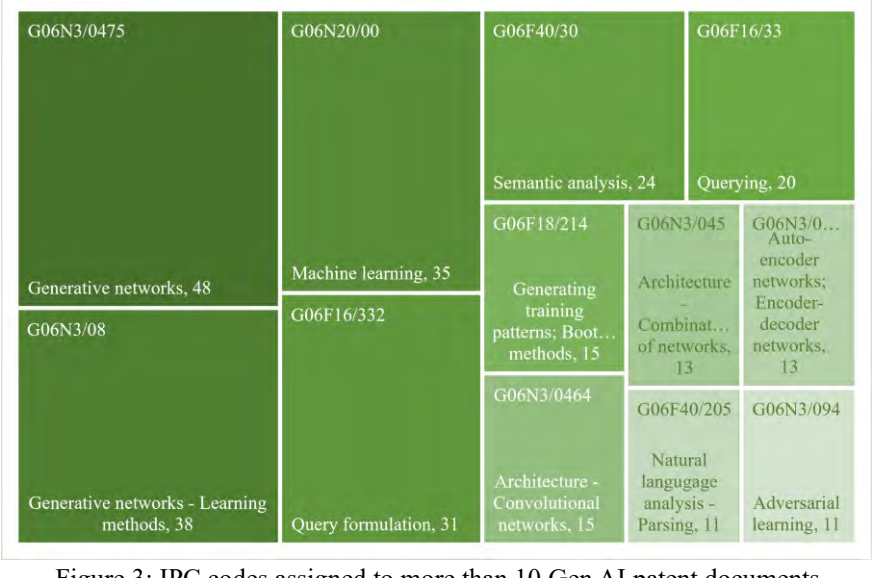


Figure 3: IPC codes assigned to more than 10 Gen AI patent documents

To detect the degree of heterogeneity of technical content related to GenAI, we have used association rule analysis. Table 4 shows eight association rules under minimal support at 1% and minimal confidence at 40%, which indicate which technical content is frequently protected together with Gen AI patent documents. Rules one to four are expected because similar content is protected together, such as query formulation and querying. However, some interesting rules exist, such as rule five and rule six. Rule number five indicates that fusion techniques for pattern recognition are registered together with feature extraction. Also, rule number six indicates that probabilistic or stochastic networks and convolutional networks are protected together with adversarial learning.

Table 6: Technical content frequently protected together with Gen AI patent documents

No.	Association rule	Support %	Confidence %
1	IF G06F16/332 - Query formulation THEN G06F16/33 - Querying	5,62	48,39
2	IF G06F16/33 - Guerying THEN G06F16/332 - Query formulation	5,62	75,00
3	IF G06F16/338 - Presentation query results THEN G06F16/33 - Querying	2,62	87,50
4	IF G06F16/338 - Presentation query results THEN G06F16/332 - Query formulation AND G06F16/33 - Quering	1,87	62,50
5	IF G06F18/25 - Fusion techniques for pattern recognition THEN G06F18/213 - Feature extraction	1,12	42,86
6	IF G06N3/0475 - Architecture-Probabilistic or stochastic network AND G06N3/0464 - Architecture - Convolutional networks THEN G06N3/094 - Adversarial learning	1,12	60,00
7	IF G06N3/08 - Learning methods AND G06F40/30 - Handling natural language data - Semantic analysis THEN G06F40/205 - Natural language analysis - Parsing	1,12	75,00
8	IF G06F16/332 - Query formulation AND G06F16/338 - Presentation query results THEN G06F16/33 - Quering	1,87	83,33

4 Conclusion

The analysis of emerging trends in Generative Artificial Intelligence (Gen AI) patents reveals a significant rise in patent applications, particularly in the last two years, with most still pending. Leading technological developments focus on real-time operations, machine learning interfaces, and AI-driven response generation, emphasising the rapid expansion of Gen AI applications across industries. The surge in Gen AI patent filings reflects the growing interest in securing intellectual property related to generative models, learning methods, and natural language processing. The WIPO document analysis [13] highlights differences in the average annual growth of patent families related to Gen AI models from 2021 to 2023: diffusion models show growth exceeding 180%, LLMs grow between 140% and 160%, autoregressive models around 40%, VAEs slightly above 30%, and GANs between 10% and 20%. While GANs patent families have experienced only moderate growth over the past three years, a similar slowdown is observed for VAE and autoregressive model patent families. In contrast, diffusion models and LLMs have demonstrated significantly higher growth rates during this period.

Notably, the majority of patent applicants are from China and the USA, with companies like Microsoft and Adobe leading the field. The growth of Chinese patent claims can be partly explained by the characteristics of the Chinese language, particularly challenges like the lack of spaces between words and the ambiguity of characters. Diffusion models are highly effective in tasks requiring controlled output and are increasingly used in sectors such as government computing, smart cities, and life sciences. Patent filings for LLMs surged from 53 in 2020 to 881 in 2023, driven by the Gen AI boom led by chatbots like ChatGPT [13]. There is still a strong need to refine search-response scenarios in Gen AI, especially for complex user queries across different languages. The rise in patents from Chinese companies and universities reflects, we believe, the demand for AI adaptations suited for users utilising Chinese characters [12].

The topics identified in Gen AI patent documents primarily address issues related to response generation and training. Notably, research indicates that adversarial attacks on GANs pose significant risks in critical fields such as healthcare and autonomous driving, where inaccurate predictions can have serious consequences [6]. Therefore, the development of robust solutions for neural networks is crucial in mitigating these risks effectively. GANs often suffer from training instability, leading to the proposal of various methods, such as network regularisation and auxiliary adversarial example regulators, to address these issues [14]. In previous research, researchers have demonstrated that techniques like gradient penalties [2] and some others enhance GANs stability and improve output quality. Thus, it can be inferred that training, as a key concept in many new-gen AI patents, emphasises a focus on innovations aimed at enhancing the stability and accuracy of AI networks.

Future research on the GenAI domain should include network analysis to discover connections between patent applicants from diverse sectors and countries, showing collaboration patterns and knowledge flow in the domain. Citation analysis should also be conducted to detect relationships between patents and explain how patents influence each other.

5 References

[1] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y. Generative Adversarial Nets. Advances in Neural Information Processing Systems 3(11), 2014.

[2] Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., Courville, A.C. Improved Training of Wasserstein GANs. Neural Information Processing Systems. DOI:10.48550/arXiv.1704.00028, 2017.

[3] Ho, J., Jain, A., Abbeel, P. Denoising diffusion probabilistic models. Advances in neural information processing systems, 33:6840-6851, 2020.

[4] Kingma, D. P., Welling, M. Auto-Encoding Variational Bayes. arXiv preprint arXiv:1312.6114, 2014.

[5] Krizhevsky, A., Sutskever, I. and Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. In: Pereira, F., Burges, C.J.C., Bottou, L. and Weinberger, K.Q., Eds., Advances in Neural Information Processing Systems, Vol. 25, Curran Associates, Inc., Red Hook, NY, 1097-1105, 2012.

[6] Malik, J., Muthalagu, R., Pawar, P. A Systematic Review of Adversarial Machine Learning Attacks, Defensive Controls and Technologies. IEEE Access. 12:99382-99421, 2024.

[7] McKinsey& Company. The Economic Potential of generative AI: The next productivity frontier. McKinsey Report, 2023.

[8] Nichol, A. Q., Dhariwal, P. Improved denoising diffusion probabilistic models. In International conference on machine learning (pp. 8162-8171), 2021.

[9] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I. Language Models are Unsupervised Multitask Learners. Technical report, Open Ai San Francisco, California, United States, 2019.

[10] Sajeeda, A., Hossain, B. M. Exploring generative adversarial networks and adversarial training. International Journal of Cognitive Computing in Engineering, 3:78-89, 2022.

[11] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ...

Polosukhin, I. Attention Is All You Need. (Nips), 2017. arXiv preprint arXiv:1706.03762, 10, S0140525X16001837, 2017.
[12] Wang, S., Sun, Y., Xiang, Y., Wu, Z., Ding, S., Gong, W., Feng, S., Shang, J., Zhao, Y., Pang, C., Liu, J., Chen, X., Lu, Y., Liu, W., Wang, X., Bai, Y., Chen, Q., Zhao, L., Li, S., Sun, P., Yu, D., Ma, Y., Tian, H., Wu, H., Wu, T., Zeng, W., Li, G., Gao, W., Wang, H. ERNIE 3.0 Titan: Exploring Larger-scale Knowledge Enhanced Pre-training for Language Understanding and Generation. ArXiv, abs/2112.12731, 2021.
[13] World Intellectual Property Organization (WIPO). Generative Artificial Intellligence. Patent Landscape Report. Geneva: WIPO. https://doi.org/10.34667/tind.49740, 2024.
[14] Yin, X., Kolouri, S., Rohde, G.K. Adversarial Example Detection and Classification with Asymmetrical Adversarial Training. arXiv: Learning, 2019.
[15] Zhu, M., Pan, P., Chen, W., Yang, Y. DM-GANs: Dynamic Memory Generative Adversarial Networks for Text-To-Image Synthesis. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 5795-5803, 2019.

Topology-Driven Feature Extraction Based on PCA Pre-Alignment: Addressing Anisotropy in High-Dimensional Data (A Case Study)

Bojan Pažek, Slavko Arh
Rudolfovo - Science and Technology Centre
Podbreznik 15, 8000 Novo mesto, Slovenia
{bojan.pazek@ferit.hr, slavko.arh@rudolfovo.eu}

Abstract. *In this paper, we focus on so-called anisotropic objects that have been digitized using non-contact inspection methods, specifically 3D-scanning. These objects are represented as point clouds or 3D structures, which we analyze using open-source Python libraries. We explore established methods and introduce a new approach using the* `trimesh` *library to improve the alignment between scanned and reference models. Specifically, we study milling cutters, where first PCA is applied for pre-alignment and then the cylindrical and threaded parts are extracted to support feature-based registration. We treat the 3D object as a topological manifold, and by carefully slicing it (similar to surgery on manifolds), we identify key features that enable more accurate comparison of scanned objects. By isolating and aligning these critical features, like the cylindrical bodies and threaded sections, we can effectively manage the challenges posed by anisotropy, which here refers to the object's varying shapes in different directions. This method enhances the accuracy and consistency of the final registration process, ensuring more reliable alignment of complex 3D structures, as our ultimate goal is to detect any differences between versions of milling cutters – such as one in perfect condition and another that is worn out (with a particular focus in this article on emphasizing the importance of detecting key features of the 3D object). Our approach advances data-driven analysis by incorporating topological data analysis (TDA), which, while often abstract and complex, plays a vital role in advancing digital transformation.*

Keywords. Anisotropic Objects, 3D-Scanning, Feature-Based Registration, PCA, Topological Manifolds, Surgery on Manifolds, Digital Transformation

1 Introduction

There are two main ways to make objects: additive manufacturing, which is often called 3D printing, and traditional manufacturing, which uses tools and machines to cut, shape, or assemble parts. Both methods have their own benefits and uses. No matter how an object is made, it is important to check that the final product matches its design in terms of size, shape, and other details. One common way to do this is by using contactless inspection with optical and laser 3D scanners [1]. This method is popular because it can quickly collect data without touching the object.

In 3D scanning, creating a reference frame is crucial because it lets us compare the scanned object to the reference model. The scan turns the object's shape into a 'point cloud', which shows the surface of the object [2]. By aligning this point cloud with the model, we can spot any differences and check if the object meets its design specifications. This alignment step is crucial for ensuring the final product works as it should, especially for parts that must fit or work with other objects. The challenge is not only capturing the surface details but also aligning them accurately with the design to make sure the part functions properly, turning the idea into a real object [3].

In this paper, we focus on anisotropic objects, particularly milling cutters. Anisotropic objects have properties that vary depending on the direction in which they are measured. Mathematically, this means that their features, such as shape or material properties, are not uniform in all directions. For example, a milling cutter have different structural and geometric characteristics along its cylindrical axis compared to its radial direction [4], [5]. Our case study aims to optimize computational methods to achieve the best possible alignment between the scanned models and the reference models of milling cutting tools.

Since this work is part of a larger project [6], here we will focus only on feature-based registration [7] by treating the 3D object as a topological manifold. A topological manifold is a mathematical space that, on a small scale, resembles Euclidean space but may have a different structure on a larger scale. In our case, this means that although the object has a complex overall shape, small parts of it (such as the cylindrical or threaded sections) can be studied and aligned as if they were simpler, flat surfaces [8], [9].

Carefully slicing the object, much like surgery on manifolds [10], [11], allows us to isolate key features and effectively address the challenges posed by the object's directional variability. This approach improves both the precision and reliability of the registration process, ensuring more accurate alignment of complex 3D models. Ultimately, the goal of the project is to detect any differences between multiple versions of milling cutters, such as comparing a new cutter with one that shows signs of wear; however, due to the concise nature of this conference paper, we focus only on a small part of the procedure, emphasizing the importance of topology, without delving into the full technical details of the process.

By presenting both the theoretical foundations and practical applications of advanced computational methods, this work seeks to make a valuable contribution to the fields of metrology, precision manufacturing, and data analysis, with a focus on milling cutters (case study). The results of our testing demonstrate the effectiveness of the proposed solutions in enhancing the accuracy and efficiency of 3D object alignment for industrial use. Furthermore, this approach strengthens data analysis through topology (TDA), a key technique that plays an essential role in advancing digital transformation.

2 From Geometry to Topology: PCA-Driven Alignment for Anisotropic Objects

Geometry and topology are closely related, yet distinct, with each focusing on different aspects of objects. While geometry focuses on the shape and size of objects, topology is concerned with how objects are connected and arranged, regardless of their precise dimensions. In our work, we transition from geometric representations, such as point clouds, to a topological approach, which provides deeper insights into the structure of complex objects like anisotropic milling cutters.

Initially, geometric data, here a point cloud, is used to capture the shape of an object. However, understanding the true nature of the object requires more than just its raw shape. Topology – in this context, the study of how an object's elements (vertices, edges, and faces) are connected – is essential for capturing its structural properties. For example, STL files, which we use as a final product in 3D scanning, describe an object's surface through a mesh of interconnected triangles. This connectedness is crucial in our case because milling cutters are anisotropic, meaning their features vary depending on direction. Analyzing these features through topology allows us to better understand and manipulate the object's structure, ensuring that we maintain the integrity of its connections even when the geometry might change due to scaling or some deformation.

To illustrate the significance of topology in mesh structures and how connectivity plays a critical role in determining the quality and consistency of meshes, let us present the following theorem.

Theorem 2.1 (Vertex Valence Uniformity in Manifold Meshes)**.** *In a uniformly connected, manifold mesh represented by an STL file, the average vertex valence* $(\bar{V})$ *is closely related to the mesh's average face degree* $(\bar{F})$*, following the relation:*

$$\bar{V} \approx \frac{2E}{V}$$

where V *is the number of vertices and* E *is the number of edges in the mesh. For a mesh composed exclusively of triangles* $(\bar{F} = 3)$*, this simplifies to* $\bar{V} \approx 6$*, assuming the mesh is large and resembles a closed surface.*

Proof. It is important to note that while many manifolds can be triangulated, there are exceptions in higher dimensions (particularly 4 and above), as demonstrated by Ciprian Manolescu in 2013 [12]. However, since we are concerned only with 2D surfaces embedded in 3D space (as is common in mesh representations such as STL files), we can safely assume that the manifolds we are working with can be triangulated. Therefore, without loss of generality, we proceed under the assumption that the mesh is triangular.

The proof is based on the Euler characteristic for a closed manifold surface. For a triangular mesh, each face has three edges, and every edge is shared by two faces. Therefore, the total number of edges E can be related to the number of faces F by: $2E = 3F$. Using Euler's polyhedron formula for a closed surface, we have:

$$V - E + F = 2 \tag{1}$$

Substituting $E = \frac{3F}{2}$ into this formula and solving for the relationship between vertices and edges, we derive the approximate value of $\bar{V}$, which is 6 for triangular meshes as F becomes large. □

This theorem states that in a well-structured manifold mesh, particularly those composed of triangles, the average number of edges connected to each vertex (vertex valence) converges toward a specific value – 6 in the case of triangular meshes, such as those in STL files. It also implies that achieving stability requires working with large meshes, which in turn demand efficient computational tools to manage. This is where PCA becomes crucial, as it allows us to handle high-dimensional data and align large, complex meshes with precision and efficiency.

Principal Component Analysis (PCA) is a statistical technique used to reduce the dimensionality of high-dimensional data while preserving its most important features. In

the context of 3D datasets, PCA is applied to identify the principal axes of inertia, which represent the directions along which the data varies the most. This technique allows us to align two objects by matching these axes, providing an initial guess for their relative orientations. In its simplest form, the principal axes of inertia are given by the eigenvectors of the covariance matrix $\mathbf{I}$ of the data:

$$\text{Principal Axes of Inertia} = \{\text{Eigenvectors of the covariance matrix } \mathbf{I}\} \tag{2}$$

where the covariance matrix $\mathbf{I}$ is defined as [13]:

$$\mathbf{I} = \sum_{i=1}^{n} \frac{1}{n} \begin{bmatrix} y_i^2 + z_i^2 & -x_i y_i & -x_i z_i \\ -x_i y_i & x_i^2 + z_i^2 & -y_i z_i \\ -x_i z_i & -y_i z_i & x_i^2 + y_i^2 \end{bmatrix} \tag{3}$$

By aligning the principal axes of inertia of two objects, we establish a rough alignment, which is crucial for the effective application of more precise iterative methods, such as the Iterative Closest Point (ICP) algorithm.

For anisotropic objects like milling cutters, where features vary depending on direction, PCA is particularly useful. It provides a method to manage these directional variations, ensuring that the alignment accommodates the complex and non-uniform nature of the object. The following images (see 1) illustrate the role of PCA in aligning anisotropic objects. The image to the left shows the original meshes with their respective principal axes of inertia, while the second image demonstrates the result after aligning the principal axes, providing a rough initial alignment of the two objects.

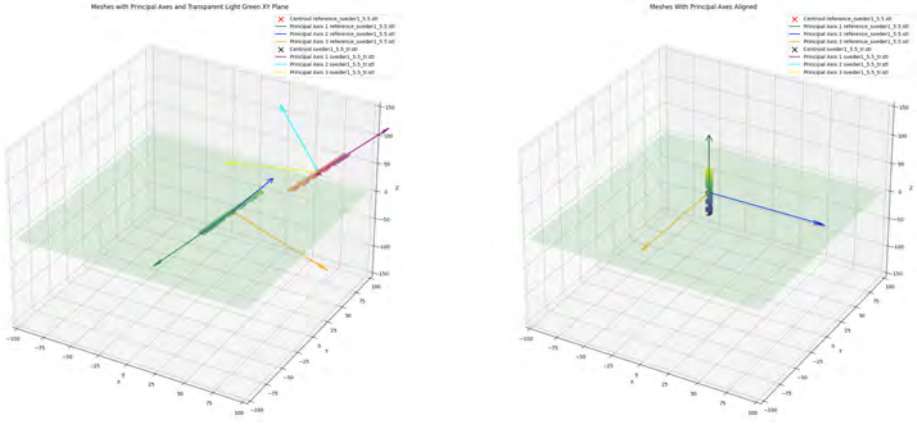


Figure 1: Comparison of Meshes Bbefore and After PCA-based Pre-alignment

The Python library `sklearn.decomposition.PCA` allows us to efficiently calculate the variance ratios, which provide insight into the anisotropy of the data. For instance, the mesh: `reference_sveder1_5.5.stl` exhibits anisotropic behavior, with variance ratios: $[0.98940233, 0.00535538, 0.0052423]$.

3 Circle Fitting and Feature Extraction in Milling Cutters

In this section, we focus on the process of circle fitting and feature extraction in milling cutters. Using Python, we successfully extracted the top section vertices and fitted circles for both meshes, providing the foundation for identifying and isolating the cylindrical and threaded parts of the cutters. The circle fitting was performed using a least squares solution.

Based on the comparison of the top cross-sections, the centers and radii of the two milling cutter meshes are nearly identical, with minimal differences. This outcome provides a decision-making point: if the cross-sections are sufficiently similar, we can proceed with the alignment process. However, if significant differences are found, it may suggest that the inspected milling cutter does not match the corresponding model in our database, prompting a search for a more suitable reference before continuing. The Figure 2 illustrates the results of this process.

```
The top cross sections are similar.
reference_sveder1_5.5.stl and sveder1_5.5_tr.stl
--------------------------------------------------
Cross-section comparison details:
center1                          [ 0.08348513 -0.08294129]
center2                          [ 0.0834852  -0.08294148]
center_difference                2.018695757564234e-07
radius1                          2.729029887996431
radius2                          2.7290298191175952
radius_difference                2.523931169502387e-08
centers_similar                  True
radii_similar                    True
```

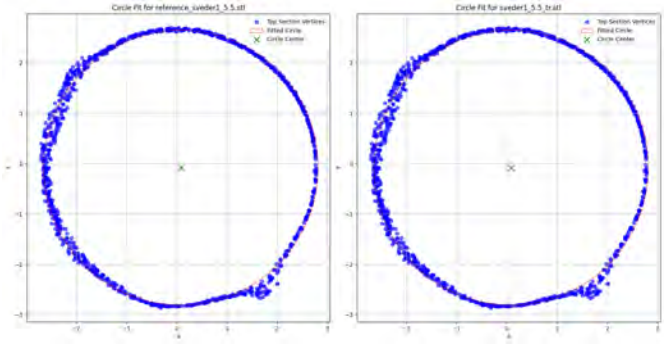


Figure 2: Comparison of Circle Fitting for Two Milling Cutter Meshes

To ensure precise alignment, we first compute the translation needed to align the centers of the two fitted circles. This is done by calculating the transformation matrix that translates the center of one mesh to match the center of the other, allowing us to apply the transformation and align the two meshes before proceeding with further analysis.

After aligning the centers of the two fitted circles, we move on to extracting the cylindrical and threaded sections of the milling cutters. The cylindrical sections were extracted

with 6,267 vertices each, and the cylindrical height was measured as approximately 56.7 units for both meshes (see Figure 3). This process, resembling surgery on a manifold, allows us to precisely extract the cylindrical part, ensuring that the feature-based registration focuses on key geometric characteristics for more reliable alignment. Notably, we did not use K-means or other advanced algorithms to extract the main features of our objects. Instead, simple topology was employed, demonstrating the power of these techniques. While advanced algorithms are often computationally expensive, topology and geometry can efficiently reveal underlying structures in the data.

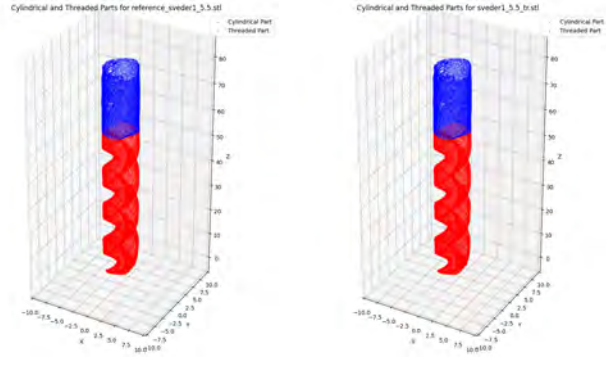


Figure 3: Extraction of the Cylindrical and Threaded Parts of Milling Cutters

The feature extraction process, including circle fitting and isolating the cylindrical and threaded sections, enables precise alignment and comparison of anisotropic milling cutters. Although the top section does not need to be perfectly circular, our method is adaptable and can fit various shapes (predefined geometric primitives like circles, polygons, etc.). Based on this, it can accurately extract the 'cylindrical' and threaded parts, ensuring that the method effectively handles the complexities of varying shapes and provides a robust solution for detecting differences between scanned and reference models, ultimately improving the reliability of feature-based registration.

4 Conclusion - Towards the Future Work

This study is a component of a broader initiative titled *Optimizing Computational Techniques for Precision Alignment of 3D Scanned Milling Cutters*. Here, we focus on feature-based registration by treating the 3D object as a topological manifold. This approach, inspired by concepts like surgery on manifolds, allows for the isolation of critical features, such as cylindrical and threaded sections, when the object exhibits anisotropic behavior. By leveraging this topological perspective, we have demonstrated an effective method for aligning scanned models.

We acknowledge the valuable feedback from the anonymous referee, which emphasized the importance of testing our approach on objects that are not entirely identical. However, this study intentionally focuses on cases where objects are nearly identical, as we aim to establish a foundation for future applications of our **Hausdorff metric**. Specifically, we seek to determine whether our implementation of this metric can detect even the slightest discrepancies. This focus is crucial, as large variations are readily detected by standard methods. The precision needed to distinguish such closely resembling objects

is essential for the ultimate goal of this project: reliably identifying variations between a new milling cutter and one showing signs of wear.

The results presented here represent only a portion of our overall research objectives. Future work will expand beyond the current focus, incorporating more complex scenarios and refining the techniques for larger-scale industrial applications. This work is a step towards ensuring that the precision and reliability of alignment methods can meet the demanding requirements of digital analysis in manufacturing.

5 Acknowledgments

This work is a continuation of the research conducted by the authors of [6], and we would like to express our gratitude to them for providing the scans of milling cutters that allowed us to implement and further develop our ideas.

References

[1] Natalia Lishchenko, Garret E. O'Donnell, Mark Culleton. *Contactless Method for Measurement of Surface Roughness Based on a Chromatic Confocal Sensor*. Machines, 2023, 11(8), 836. DOI: 10.3390/machines11080836.

[2] Ulrich Stenz, Jens Hartmann, Jens-André Paffenholz, Ingo Neumann. *High-Precision 3D Object Capturing with Static and Kinematic Terrestrial Laser Scanning in Industrial Applications—Approaches of Quality Assessment*. Remote Sens., 2020, 12(2), 290. DOI: 10.3390/rs12020290.

[3] Jang, KH., Yoon, HS., Lee, HT. et al. *50 nm Scale Alignment Method for Hybrid Manufacturing*. Springer, 2023., Int. J. Precis. Eng. Manuf. 21, 2407–2417 (2020) DOI: 10.1007/s12541-012-0026-2.

[4] Libretexts, Introduction to Anisotropy, `https://eng.libretexts.org/Bookshelves/Materials_Science/TLP_Library_I/02%3A_Introduction_to_Anisotropy`, downloaded: September 13^{th} 2024.

[5] G.K. Larsen. *Characterization of Anisotropic and Shape-Selective Nanomaterials: Methods and Challenges*. In: Hunyadi Murph, S., Larsen, G., Coopersmith, K. (eds) *Anisotropic and Shape-Selective Nanomaterials*. Nanostructure Science and Technology, Springer, Cham, 2017. DOI: 10.1007/978-3-319-59662-4_4.

[6] Jožica Piškura, Marko Rudolf, Simon Muhič, Janez Povh. *Visokozmogljivo računanje optimalnega prileganja 3D objektov*. Received: 7.11.2022, Revised: 8.11.2022, Accepted: 8.11.2022. *Svet strojništva*, ISSN-2350-3505, 2022, Letnik 11, št. 03/06.

[7] Uroš Urbas, Timotej Hrga, Janez Povh, and Nikola Vukašinović, *Novel alignment method for optical 3D gear metrology of spur gears with a plain borehole*, Measurement, Volume 192, 2022, 110839, DOI:10.1016/j.measurement.2022.110839.

[8] Zongwei Yao, Quanxiao Zhao, Xuefei Li, Qiushi Bi, *Point cloud registration algorithm based on curvature feature similarity*, Measurement, Volume 177, 2021, https://doi.org/10.1016/j.measurement.2021.109274.

[9] Weimin Li, Pengfei Song, *A modified ICP algorithm based on dynamic adjustment factor for registration of point cloud and CAD model*, Pattern Recognition Letters, Volume 65, 2015, Pages 88-94, https://doi.org/10.1016/j.patrec.2015.07.019.

[10] C.T.C. Wall. *Surgery on Compact Manifolds*. American Mathematical Society, 1999.

[11] W. Lück and T. Macko. *Surgery Theory: Foundations*. Springer, 2024.

[12] Ciprian Manolescu, *Lectures on the triangulation conjecture*, arXiv preprint arXiv:1607.08163, 2016. Available at https://arxiv.org/abs/1607.08163

[13] Kuttler, K. *Orthogonality*, in A First Course in Linear Algebra, LibreTexts, 2007. Available at: https://math.libretexts.org/Bookshelves/Linear_Algebra/A_First_Course_in_Linear_Algebra_(Kuttler)/07%3A_Spectral_Theory/7.04%3A_Orthogonality, downloaded: September 14^{th} 2024.

MODELLING BORDER CROSSING IN IRREGULAR MIGRATIONS ON THE WESTERN BALKANS ROUTE

Aljaž Blatnik[1], Blaž Rodič[1,2]
[1]Faculty of Information Studies
Ljubljanska cesta 31A, 8000 Novo mesto, Slovenia
[2]Rudolfovo - znanstveno in tehnološko središče Novo mesto
Podbreznik 15, 8000 Novo mesto, Slovenia
{aljaz.blatnik, blaz.rodic}@fis.unm.si

Abstract: *In this contribution, we construct a simulation model that integrates several modelling methodologies. This model serves to simulate the dynamics of migration from source countries to the European Union via a series of border crossings on the Western Balkans route. By incorporating Agent-Based, System Dynamics and Discrete Event Simulation methodologies, we aim to comprehensively capture the complexities of migration processes. Our model provides a foundational framework for future development and enhancements in understanding migration dynamics. It enables us to explore various scenarios, assess the impacts of different factors, and potentially inform policymaking and intervention strategies related to migration management.*

Key Words*: Simulation modelling, Agent-based modelling, Discrete event Simulation, Migrations, Refugees*

1 Introduction

Human migration is a significant research topic with notable economic effects [1]. International migration has steadily increased since WWII. In 2020, about 281 million people lived outside their birth country for a year or more, representing only 3.6% of the global population. Europe hosts the largest migrant population at 87 million, followed by North America (59 million) and Northern Africa and Western Asia (50 million). While most international migrants reside in high-income countries, South-South migration has grown, with 36% of global migrants, or 82 million, residing in the Global South [2].

But not all migration is voluntary. In any given year, tens of millions of people move to escape political violence, war, hunger, deprivation, and the vagaries of climate change, becoming refugees, asylum seekers, or internally displaced persons. At the end of 2020, the number of »persons of concern« to the United Nations High Commissioner for Refugees (UNHCR) was 82.4 million (1 percent of world's population) [3][4].

Migration, migration flows, causes and their consequences are explained by various migration theories, which range in the fields of sociology, economics, anthropology, geography, demography, history, law, political sciences [2], and we can also add psychology. The factors influencing migration decisions and the choice of destination countries are extensively documented in numerous publications [5][6][7][8]. Further

literature describes research on regional migration systems [9][10][11], the impact of conflict on migration [12][13] and even extreme forms of violence such as genocide [14], security aspects of migrations [15][16], economic factors [17], and public opinion in the destination countries [18].

Previous studies on human migration generally make simplistic assumptions about the decision model of migration, or do not consider fluctuations in birth/death rates, age distributions as well as networks ties between countries. Because of that it is important to develop an agent-based model with recent economic trends, birth policy changes and climate trends in mind [19].

In our research we have focused on the Western Balkans route, which refers to the irregular arrivals in the EU through the region: Albania, Bosnia and Herzegovina, Kosovo, Montenegro, North Macedonia and Serbia [20]. Our model integrates several simulation modelling methodologies: Agent-Based Modelling, Discrete Event Simulation and System Dynamics. We portray each refugee as an individual agent embarking on a journey from their home country to the border of the European Union. Throughout this journey, agents navigate a series of decisions influenced by both individual preferences and external factors, such as weather conditions and border policies across various countries. This combination allows for a comprehensive representation of migration dynamics, capturing the interplay between personal choices and broader environmental and policy influences. The main focus of the presented model is the process that takes place at each border that agents (refugees) take on the journey to the European Union.

2 Methodology

Our main simulation modelling tool is AnyLogic, which is a multimethod simulation software tool that allows us to integrate all of the selected methodologies within our model. It's used for simulating complex systems and processes across various industries such as manufacturing, logistics, healthcare, transportation, and more [21]. The idea of multimethod modeling is simple: to seamlessly integrate different methods of modeling and simulation to overcome the drawbacks of individual approaches and get the most from each one. Combining different methods leads to efficient and manageable models without using workarounds. In our model, we have used all three major methodologies used to build dynamic simulation models:

- System dynamics (SD): assumes a high abstraction level and is primarily used for strategic level problems, such as market adoption rates and social process dependency;
- Discrete event modeling (DES): mainly used at operational and tactical levels, like manufacturing processes and equipment investment evaluation;
- Agent-based modeling (ABM): is used at all levels, with the agents possibly being any active entity, example applications include supply chain optimization and epidemiology [22].

Models that address the migration in general have a long history [23], however models that explore forced migration (such as refugee movements) specifically are less common [24][25]. Forced migration is challenging to predict and model due to the extreme push-pull dynamics involved, such as war and other types of conflict. These dynamics significantly influence population movements [26]. Only recently the forced migration

has been studied through computational modeling, but majority of these models are agent-based, which focus on individual behavior and movements of refugees [27].

3 Border Crossing Model

The core concept of the model involves agents (representing refugees) moving through a predefined spatial environment. Each agent starts their journey at an initial location, representing the refugee's country of origin. The arrival rate (into the model) of these agents is determined by a population model that simulates the flow of individuals over time. Upon initiation, each agent's goal is to reach the EU by navigating through various border checkpoints, encountering both legal and illegal pathways. The agents are aware of their target destination country at the beginning, however the target country may change during the travel. An agent's destination and chosen route can shift dynamically in response to changing conditions, such as a country closing its borders, as well as factors like route accessibility, cost, and difficulty. As they move through physical space, agents' travel speeds are influenced by factors like fatigue levels, weather conditions, and access to transportation. Countries along the route each have unique legal frameworks that regulate border accessibility and influence the journey's difficulty. These national policies, combined with the geographical distances agents must cover, impact how quickly and efficiently agents can progress toward the EU. As border regulations evolve in response to refugee situations, the journey may become more challenging or accessible for agents moving along the route.

Our model's core objective is to be dynamic, enabling seamless additions or removals of source countries and border crossings as deemed necessary along the journey. The model consists of several core elements.

3.1 Population model for countries

In our model, countries determine the arrival rate of refugee (agents). Each country is characterized by its own birth and refugee rates. The birth rate dictates how the population of the country increases over time, while the refugee rate specifies the percentage of the population that decides to leave the country as refugees. Both the birth rate and refugee rate may change over time, depending on the country's level of stability.

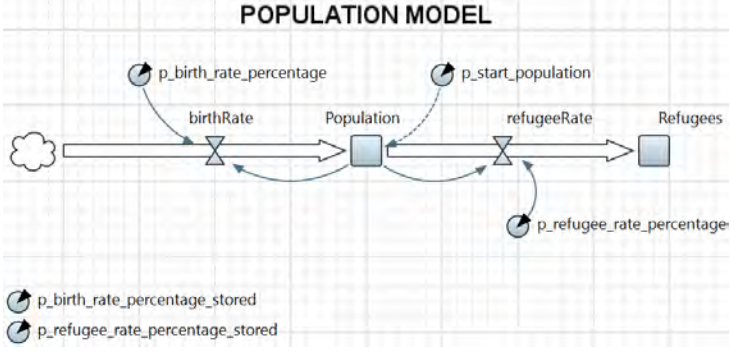


Figure 1: Population model

In Figure 1, we observe the system dynamics model used for population calculation in each country. Currently, the model includes only one such representation. The model comprises two stock elements: Population and Refugees. The Population stock is initialized with the p_start_population parameter, set to the actual population of the desired source country (in our case, Afghanistan). The initial value of the Refugees stock is set to 0. We also define the values for both p_birth_rate_percentage and p_refugee_rate_percentage. The model consists of two flows. The first flow, birthRate, simulates population growth, depending on the birth rate percentage and the total population of the country. The second flow, refugeeRate, simulates the departure of refugees from the country, depending on the refugee rate percentage and the total population of the country.

3.2 Stability model

Our population model is contingent upon the stability of the country. The system dynamics model used to calculate stability is depicted in Figure 2 below.

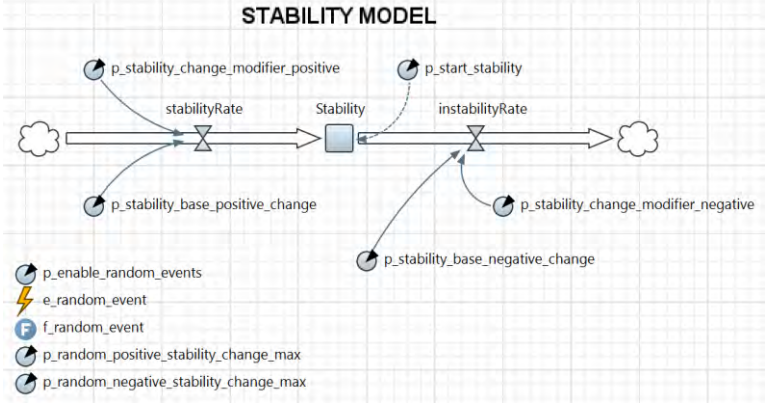


Figure 2: Stability model

The primary component of the model is the Stability stock element, which represents the current level of stability within the country. Stability increases are determined by the inflow parameter, stabilityRate, while decreases are determined by the outflow parameter, instabilityRate. The initial value of Stability is set by the parameter p_start_stability, which is initialized to 0. Both the stabilityRate and instabilityRate parameters change dynamically over time, but their base values remain constant. The remaining two parameters, p_stability_change_modifier_positive and p_stability_change_modifier_negative, undergo random fluctuations on a monthly basis when random events are enabled (specified by setting the parameter p_enable_random_events to true). The Event element e_random_event governs this process, ensuring that the random change occurs each month. The modifications to the mentioned stability parameters are done in function f_random_event.

3.3 Refugees

The core components of our model are refugees, represented as agents navigating through a discrete event simulation. Each agent symbolizes a group of refugees, which helps mitigate potential performance issues that could arise from simulating individual refugees, especially at higher arrival rates.

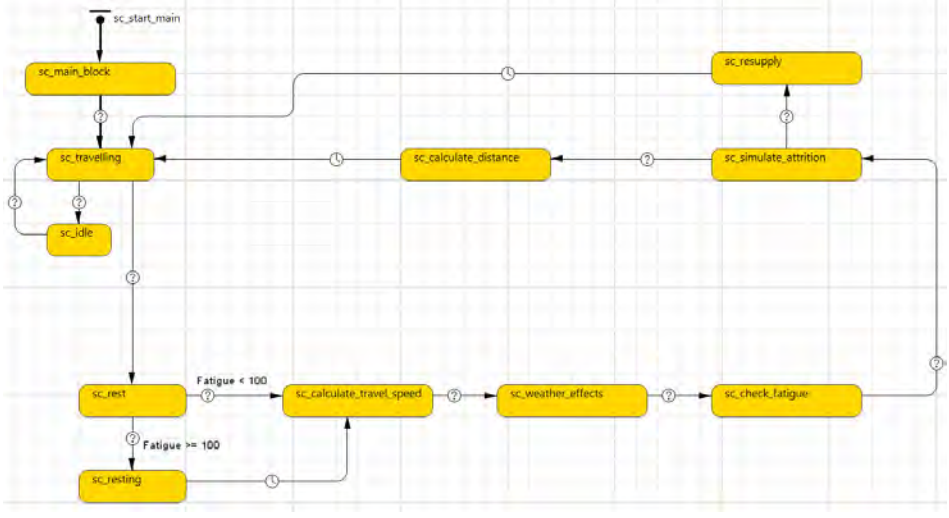


Figure 3: Refugee_Group agent - main Statechart

Figure 3 illustrates the primary Statechart for the refugee agents. This Statechart is active whenever a group of refugees is traveling; otherwise, they remain in the sc_idle state until movement resumes. The sc_main_block (right after Statechart entry point sc_start_main), the initial block of the Statechart, currently randomizes a few parameters. The block utilizes the function f_init_parameters to assign values to the parameters, providing flexibility for future expansion if needed. The next block in the Statechart is the sc_travelling block. The transition from sc_main_block occurs when the parameter tracking the refugee group's movement is set to true, indicating the group has started moving. The sc_travelling block tracks the group's progress and marks the beginning of the movement cycle. To monitor the journey, we use the parameter p_km_to_travel, which tracks the remaining distance in kilometers. This parameter, along with several others, is initialized upon the agent's arrival through a function in the Main agent (the agent with the discrete event model).

Following the sc_travelling block in the Refugee_Group agent, we encounter the sc_rest block. This block serves as the trigger for two transitions, checking whether the agents' fatigue level has reached 100 or higher. If fatigue exceeds this threshold, the agent halts all movement for a few days to allow the fatigue level to decrease. The duration of rest days is determined randomly. Upon completion of the rest period, the agent proceeds to the sc_calculate_travel_speed block. Alternatively, within the sc_calculate_travel_speed block, the agent resumes its journey if its fatigue is less than 100, proceeding without any delay.

In the subsequent phase of the Refugee_Group agent Statechart, we evaluate weather and fatigue effects (refer to Figure 3). Weather conditions are simplified to two states: good and bad. In adverse weather, travel speed decreases, while good weather permits optimal travel speed with no hindrance. The weather parameter (reference) is adjusted in the Main agent through the Travelling delay block within the discrete event model on a daily basis, updating whenever the Refugee_Group agent enters the block.

The second stage of the effects calculation involves determining the impact of fatigue on travel speed. Depending on the fatigue level, the travel speed may decrease, as illustrated in Table 1 below.

Table 1: Fatigue impact on travelling speed

FATIGUE LEVEL	FATIGUE IMPACT
<= 50	No impact
<= 75	50% travelling speed
<= 100	25% travelling speed
> 100	0% travelling speed

As fatigue increases, the corresponding travel speed decreases. When fatigue surpasses a level of 50, the speed reduces to 50%. Upon exceeding a fatigue level of 75, the travel speed drops further to 25%. Beyond a fatigue level of 100, the group ceases to move altogether.

The final step in the main Statechart of the Refugee_Group agent involves simulating attrition and calculating the remaining distance to the goal. Initially, we randomly determine the duration of the group's travel, ranging from a minimum of 1 hour to a maximum of 24 hours. The level of fatigue incurred by the refugees depends on the duration of travel; longer journeys without breaks result in higher fatigue levels. Subsequently, we calculate the distance covered, factoring in the current speed of the group and the hours of travel. The resulting distance is then deducted from the parameter p_km_to_travel, representing the remaining kilometers to reach the goal (the next point in the discrete event model within the Main agent). Another process within the agent involves the potential for the refugee group to either gain or lose members. Along their journey, stationary refugees may decide to join the group and attempt the border crossing together. Conversely, some existing members of the refugee group might find their current location satisfactory and choose to leave the group. The possibility of the group gaining or losing a member can occur every 5 days, although it is not guaranteed. In the Main agent we include all of the described agents and models. The framework that helps agents move around is discrete event system.

4 Results

During the final phase of testing and running the model, we opted for a Monte Carlo experiment. In this experiment, all values were set to default except for the initial stability, refugee rate percentage, and birth rate percentage. Uniform functions were utilized to assign random values to these parameters. The experiment consisted of 50 runs, each spanning 365 days.

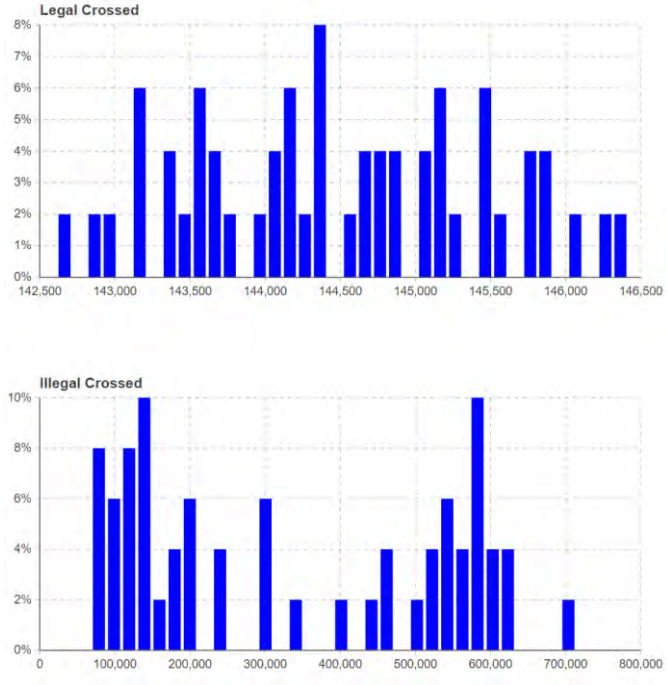


Figure 4: Results of Monte Carlo experiment

In Figure 4, we present the outcomes of the Monte Carlo experiment. The lowest recorded count of legal crossings stands at 142.615, while the highest count reached 146.415. Concurrently, illegal crossings range from a minimum of 68.655 to a maximum of 713.775. Based on these findings, we deduce that our current border crossing system is capable of legally processing from 142.000 to 147.000 refugees yearly. Additionally, it is evident that the count of illegal crossings can vary, sometimes falling below the number of legal crossings at around 130.000 illegal crossings, while at other times surging significantly higher, especially around the 580.000 mark. Furthermore, we notice a distinct trend in the distribution of illegal crossings, which tend to cluster around two extremes. Either the count of illegal crossings remains relatively low, with values concentrated around 130.000, or it spikes considerably higher, forming a cluster around 580.000. In contrast, legal crossings exhibit a more uniform distribution across the range of values. Throughout all the runs, the average count of legal crossings hovers around 144.535, whereas the average count of illegal crossings across all runs stands at 345.250. This indicates that, on average, the number of illegal crossings will consistently surpass that of legal crossings.

5 Conclusion

We successfully developed the AnyLogic model to simulate the migration from any country into Europe. The model incorporates the decision-making processes of the refugees (agents) and the status of the countries through which the agents are travelling.

The current version of model has several limitations: the agent decision model logic and parameters are based on educated guessing, and further study of previous research is needed to develop a model that will more closely correspond to the real-life system. The model is not yet calibrated or validated, and thus cannot be used to explain past events or predict future developments on the Western Balkans route.

However the current model serves as the foundation for a more sophisticated version we intend to develop in the future. Our plan involves upgrading the border crossing process by integrating additional real-life data to more accurately reflect the conditions of the countries along the route to the European Union. We aim to fine-tune the probabilities and also account for both government and non-government workers to better align with the actual circumstances (chances that authorities capture illegal crossings, the speed at which legal crossings can be processed, etc.). Presently, we are in the process of gathering data to enrich the model and implement further enhancements.

There is a lot of options left open to further improve the model and there is also space for additional source countries and border crossings in the process, meaning the goal to create a very flexible model was achieved. We will expand the current agents to better represent their decisions during the whole border crossing process and also add additional agents, which may represent authorities, such at police and their ability to successfully implement government decisions and capture illegal refugees. In the future we might also add visual presentation where the agent will be able to move across actual map, giving us more insight to where the most important refugee route is and how it changes with time.

Our main future task is to gather more data to both plan for the future upgrades and to compare the results we get from the model with the real-life data. The data will also allow us to calibrate and validate future versions of the model, allowing the use of the model for prediction of migration flows.

8 Acknowledgements

The work was supported by the Slovenian Research and Innovation Agency [Research program No. P1-0383, Complex networks and research projects J5-4585: “MIG@B, Modelling of international migrant flows in the Balkans” and J7-3156: “Synergistic Integration of Quantitative Sociology and STEM fields to resolve Critical Social Dilemma: the Cases of Vaccination, Migration, and Corruption”].

9 References

[1] OECD. (2014). *Is migration good for the economy?* Retrieved from https://www.oecd.org/migration/OECD%20Migration%20Policy%20Debates%20Numero%202.pdf (16. 9. 2024).
[2] Brettell, C. B. & Hollifield, J. F. (2023). *Migration Theory: Talking across Disciplines* (4th ed.). New York, NY: Routledge.
[3] Hazan, M. (2021). International Migration and Refugee Movements in Latin *America*. Hollifield, J. F. & Foley, N. (ur.), *Understanding Global Migration* (pp. 312 – 341). Stanford, CA: Stanford University Press.
[4] Hollifield, J. F. (2021). *Driven Out*. Retrieved from https://www.wilsonquarterly.com/quarterly/_/driven-out (17. 9. 2024).
[5] Mayda, A. M. (2010). International migration: A panel data analysis of the

determinants of bilateral flows. *Journal of Population Economics*, 23(4), 1249–1274.
[6] Mayda, A. M., & Patel, K. (2005). International Migration Flows: the Role of Destination Countries ' Migration Policies. *In Policy*.
[7] Baban, F., Ilcan, S., & Rygiel, K. (2017). Syrian refugees in Turkey: pathways to precarity, differential inclusion, and negotiated citizenship rights. *Journal of Ethnic and Migration Studies*, 43(1), 41–57.
[8] Düvell, F. (2019). The 'Great Migration' of summer 2015: analysing the assemblage of key drivers in Turkey. *Journal of Ethnic and Migration Studie*s, 45(12), 2227–2240.
[9] Haas, H. de. (2007). North African Migration Systems : Evolution , Transformations and Development Linkages. In S. Castles & R. D. Wise (Eds.), *Migration and Development: Perspectives from the South* (pp. 143–174). Retrieved from https://www.semanticscholar.org/paper/North-African-migration-systems%3A-evolution%2C-and-Haas/d47b442e58489769f755e478045d916622874ed9.
[10] Hein, de H. (2011). Mediterranean migration futures: Patterns, drivers and scenarios. *Global Environmental Change*, 21, Supple, S59–S69.
[11] Milan, C. (2019). Refugees at the gates of the EU: Civic initiatives and grassroots responses to the refugee crisis along the western Balkans route. *Journal of Balkan and Near Eastern Studies*, 21(1), 43–60.
[12] Justino, P. (2009). Poverty and violent conflict: A micro-level perspective on the causes and duration of warfare. *Journal of Peace Research*.
[13] Nyaoro, D. (2019). Refugee Hosting and Conflict Resolution: Opportunities for Diplomatic Interventions and Buffeting Regional Hegemons. *Advances in African Economic, Social and Political Development*, 17–32.
[14] Bučar-Ručman, A. (2020). Genocide and forced migration : the dual victimisation of refugees escaping war and genocide. *Genocide and Victimology*, 1st ed., Y. Eski, Ed. Routledge, 169–188.
[15] Kicinger, A. (2004). International migration as a non-traditional security threat and the EU responses to this phenomenon. CEFMR Working Paper 2/2004.
[16] Kopač, E. (2007). "Migracijsko ogrožanje nacionalne varnosti," in Model celovitega ocenjevanja ogrožanja nacionalne varnosti Republike Slovenije (I. Prezelj, L. Jelušič, A. Anžič, & Schwarz), Eds. Ljubljana: Ministrstvo za obrambo, Direktorat za obrambne zadeve, Sektor za civilno obrambo, 191–202.
[17] Chiswick, B. R., & Miller, P. W. (2014). Handbook of the economics of international migration. Volume 1B, The impact and regional studies.
[18] Mayda, A. M. (2006). Who is against immigration? A cross-country investigation of individual attitudes toward immigrants. *Review of Economics and Statistics*, 88(3), pp. 510–530.
[19] Lin, L., Carley, K. M. & Cheng, S. (2016). An agent-based approach to human migration movement. *Proceedings of the Winter Simulation Conference 2016*. Research Collection School Of Information Systems.
[20] European Council. (2024). *Western Balkans Route*. Retrieved from https://www.consilium.europa.eu/en/policies/eu-migration-policy/western-balkans-route/.
[21] Anylogic. (n. d. a). *Simulation Software for Every Business Challenge*. Retrieved from https://www.anylogic.com/features/ (18. 9. 2024).
[22] Anylogic. (n. d. b). *Multimethod Simulation Modeling*. Retrieved from https://www.anylogic.com/use-of-simulation/multimethod-modeling/ (18. 9. 2024).
[23] Stillwell, J. (1978). Interzonal migration: some historical tests of spatial-interaction models. *Environment and Planning A*, vol. 10 (10), pp. 1187 – 1200.

[24] Djamengo, M. D. P. & Fanokoa, P. S. (2015). System Dynamics Modeling of Impacts of Central African Republic Refugees in Eastern Region of Cameroon. *Proceedings of the 33rd International Conference of the System Dynamics Society*, pp. 19 – 23, Cambridge, MA, International Conference of the System Dynamics Society.
[25] Taylor, I. W. & Araujo, P. (2021). Using System Dynamics to Examine Alternative Futures for the Syrian Refugee Crisis. *Proceedings of the 2021 System Dynamics Conference*. Chicago, IL.
[26] Curry, T., Croitoru, A. & Crooks, A. (2023). Modeling Forced Migration: A System Dynamic Approach. 2023 *Annual Modeling and Simulation Conference (ANNSIM)*. Hamilton, ON, Canada.
[27] Suleimenova, D. & Groen, D. (2020). How policy decisions affect refugee journeys in South Sudan: a study using automated ensemble simulations. *Journal of Artificial Societies and Social Simulation,* vol. 23 (1), pp. 2.

Computer Vision for Time Series Classification

Jelena Joksimović[†‡]
[†]Faculty of Information studies
Ljubljanska cesta 31a, 8000 Novo mesto, Slovenia
[‡]Rudolfovo Science and Technology Center
Podbreznik 15, 8000 Novo mesto, Slovenia
{jelena.joksimovic}@rudolfovo.eu

Abstract. *Time series analysis is essential for uncovering patterns in large, temporally ordered datasets, such as financial transactions, and detecting important events. Traditionally, time series clustering is used when dealing with unlabeled data, relying on numerical measures of similarity to group similar sequences. However, these methods often miss visual patterns that could offer more intuitive insights. In this work, we initially applied clustering to our time series data due to the absence of labels, aiming to identify meaningful groupings based on underlying similarities. Our approach was tested on a dataset of almost 250,000 public-to-private sector financial transactions in Slovenia, covering the period from 2003 to 2020. A key challenge was identifying "suspicious" transactions, which exhibited significant changes in income corresponding to government transitions. Without labeled data, we explored clustering methods but found that the visual patterns provided by domain experts offered clearer insights. This led us to transition from clustering to a classification task by converting the time series into images. Using expert-driven rules, we generated labeled target images and assigned them to 2 classes. We artificially created over 200 images for each class and trained a Convolutional Neural Network (CNN) to classify the images based on these patterns. The CNN was then used to classify new time series images. Our results demonstrate that CNN-based image classification, guided by domain expert insights, was highly effective for detecting important patterns in our time series data, providing a more efficient approach than traditional clustering.*

Keywords. public spending, time series analysis, convolutional neural networks, computer vision

1 Introduction

When dealing with unlabeled time series data, the challenge often lies in identifying groups of similar time series, a task commonly addressed through clustering. Clustering aims to uncover underlying structures by grouping time series that exhibit similar patterns. Two main approaches to time series clustering have been developed: the raw-data-based approach and the feature-based approach [5, 3, 6, 7]. In the raw-data-based approach, which we employ in this work, algorithms operate directly on the raw time series, with the key challenge being the selection of a similarity or distance measure that

captures the temporal nature of the data. The k-Means algorithm, using Euclidean distance, was chosen as the clustering method, assigning each time series to one of several clusters, thereby revealing meaningful groupings within the dataset.
In contrast, computer vision analyzes spatial relationships in visual data, typically represented as images or videos, across multiple dimensions such as width and height. Unlike time series data, which is governed by temporal order, images are characterized by spatial patterns such as edges, textures, and shapes.
However, even thought time series are sequences of numerical values, the way how they actually **look** often plays significant role in deciding where do they belong (e.g. to which cluster), what methods to apply to them etc. This raises an intriguing question: **Can time series data be transformed into images and then analyzed using computer vision techniques?** In our use-case, by converting time series into visual representations, such as geometric line plots, we applied Convolutional Neural Network (CNN)-based image classification techniques to try to answer this question. In our approach, we leveraged domain expert knowledge[1] to generate labeled images from time series data, thus bypassing the need for traditional clustering methods. Instead, we artificially created target images based on patterns identified by experts, such as periods of inactivity or irregular financial behaviors.
By converting time series into images, we utilized CNNs to classify these visual representations into predefined categories, enabling the detection of significant patterns. This cross-disciplinary approach bridges the gap between temporal and spatial data analysis, demonstrating that image-based classification provides an effective alternative for analyzing time series data, especially when expert-driven labeling is available.

1.1 Data

In our case study, we analyze public-to-private sector financial transactions represented as time series, each with 209 data points. The primary challenge was distinguishing between "suspicious" and "regular" transactions, a task undertaken in collaboration with the Commission for the Prevention of Corruption in Slovenia (KPK). Without labeled data, we relied on domain expert insights, which indicated that transactions showing a significant increase in income during specific government periods, followed by a decline in the preceding or subsequent periods, could be considered suspicious. Conversely, transactions exhibiting no significant variations linked to political transitions were classified as regular.
Visual representations of these "suspicious" patterns (see Figure 1) that to KPK presented main starting point to investigate these relationships between public and private contracts, guided our exploration of computer vision techniques for this task. From a dataset of almost 250,000 transactions, each spanning 209 months, we first transformed their time series formats into geometric line plots (see Figure 2), that depict total transaction amounts across various government periods. These visual cues became the foundation for our approach, where we sought to classify the time series based on how the trends appear visually rather than relying solely on numerical data [2].

The original dataset includes all public-to-private sector transactions in Slovenia between January 2003 and May 2020, covering a total of 248,989 companies. Each record

[1] By experts here we mean professionals from the Commission for the Prevention of Corruption in Slovenia (KPK) with whom we collaborated on this project

[2] Later in our approach, we didn't use these particular transformations, but the time series image as is (after standardization), the idea here was to introduce the main motivation for this paper.

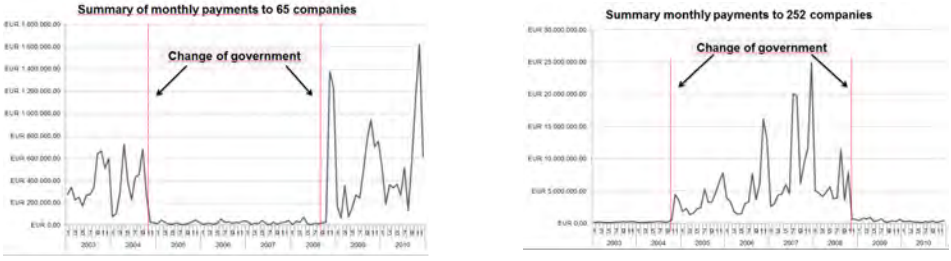


Figure 1: Illustration depicting two distinct sets of companies: The first group (on the left) experiences a decline in public sector contracts during one government's tenure, while the second group (on the right) secures new business opportunities. These figures are obtained directly from Commission for the Prevention of Corruption in Slovenia, and their methods are unknown to us.

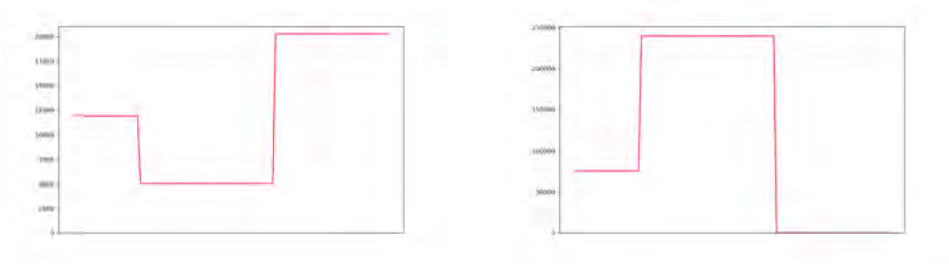


Figure 2: Two distinct geometry patterns (images to cluster) of time series after the transformation.

contains monthly income (in euros) derived from public sources, such as national and regional governments, ministries, educational institutions, and hospitals. The dataset lacks contextual details about the companies, such as ownership, contracts, or partners. In previous work [2, 4], we performed an initial filtering of the dataset, excluding companies with a total maximum income below 1 000 000 EUR.

Additionally, we filtered the time series horizontally, focusing on the period between January 2003 and February 2012. This 110-month window includes the transition periods between three key government tenures: *Rop*, *Janša*, and *Pahor*. This period was chosen due to its political significance, with government transitions likely influencing financial transactions. The table of significant events during these tenures is available in [4, 1].

Finally, a random subset of 1,000 time series was selected for further analysis. These time series were converted into images (see Figure 3) and analyzed using the methods described in the following sections. Our ultimate goal is to determine whether visual-based classification techniques, specifically using CNNs, can improve the identification of suspicious transactions, enhancing both the speed and accuracy of detecting irregular financial behavior.

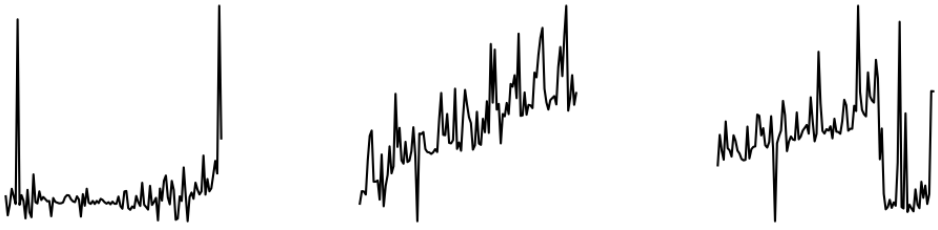


Figure 3: Three images examples of transactions.

2 Methods

To cluster the raw time series data, we will apply the k-Means algorithm using Euclidean, as an appropriate distance metric to capture the temporal dynamics of the data. After several experiments, we set the number of clusters to be $k = 12$ as the most optimal for different patterns preview. This baseline approach will allow us to compare the effectiveness of clustering based on raw numerical values versus classification based on visual representations.

The time series data, representing monthly records over a period of 110 months, was first transformed into visual representations. Each time series was plotted as a simple line chart, with no axis labels or ticks to avoid any bias during the image-based classification. This step resulted in a dataset of 1000 images for prediction (3 examples are shown in Figure 3).

To train a CNN, we artificially generated a dataset of 350 images (150 images for *Class 0 - regular* and 200 images for *Class 1 - suspicious*), where each image was of size 256x256 pixels. The goal was to leverage the visual differences between the two types of time series patterns for classification. The first part of the process involved generating artificial time series data and transforming it into images. For *Class 1*, we created 100 time series that contained zeros between the 23rd and 70th months, with random values before and after this interval, and 100 time series with the exactly opposite dynamic. This reflects a specific pattern of inactivity in the time series. In contrast, *Class 0* consisted of time series that had random values throughout the 110 months, simulating continuous activity with no intervals of inactivity [3]. Additionally, further variation within *Class 0* and *Class 1* was created by generating slightly different patterns, ensuring robustness in the training data.

After generating the time series images, a CNN was implemented to classify these images. A simple CNN architecture was used, consisting of convolutional layers for feature extraction, followed by max-pooling layers to reduce the spatial dimensions and prevent overfitting. The CNN was fine-tuned to identify patterns specific to *Class 0* (continuous activity) and *Class 1* (zero values in specified intervals). The model was trained on the generated images using binary cross-entropy loss as the classification task involved two classes (0 and 1). The dataset was split into training and validation sets using an 80-20 split, with the images being rescaled and normalized prior to training. Once the CNN was trained, it was used to classify a new set of our initial 1,000 images.

[3] Although hardcoding the distinction between zero and active months may initially appear simplistic, in our specific use case, where we have precise knowledge of the event we are targeting and the expected timing of changes in transaction dynamics, it is a well-justified approach. This method allows us to effectively capture the patterns associated with suspicious transactions, aligned with the known political transitions.

3 Results

The results of the k-Means clustering on the raw time series are shown in Figure 4. We observe that the 1,000 transactions are distributed almost evenly across the 12 clusters, making it challenging for experts to visually identify a cluster specifically associated with government transitions, as illustrated in Figure 1. In contrast, the computer vision techniques, using the CNN model, successfully classified the images, identifying only 13 out of the 1,000 transactions as *Class 1* (suspicious). This is an excellent outcome, as detecting a small subset of potentially suspicious companies that can be manually reviewed by domain experts is more practical and valuable in this context. Examples of companies classified as *Class 1* are shown in Figure 5, where it is clear that these companies align with the KPK's subset of suspicious companies, as seen in the right panel of Figure 1.

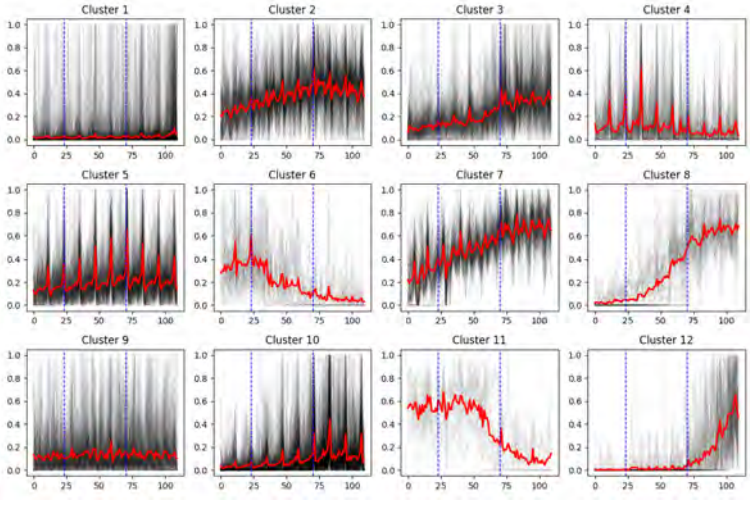


Figure 4: 12 clusters as the result of clustering the 1,000 raw time series with usual time series clustering method.

4 Conclusion and future work

In this study, we explored a novel approach to time series classification by converting time series data into images and applying CNNs for pattern detection. This method, driven by domain expert knowledge, allowed us to bypass traditional clustering approaches and instead focus on classifying time series data using visual features. By transforming time series into geometric line plots, we created two classes of images: one representing suspicious patterns with zeros in specific intervals and another representing continuous patterns in regards to the government transitions. For future work, we aim to extend this approach by incorporating semi-supervised learning. After the initial classification of time series images, domain experts will review the results, verifying and labeling the identified companies as either "suspicious" or "regular" based on additional insights not captured in the data. These expert-labeled companies will then be fed back into the model to further refine the classification process, transforming it into a semi-supervised learning task. This feedback loop will enhance the model's accuracy by integrating human expertise,

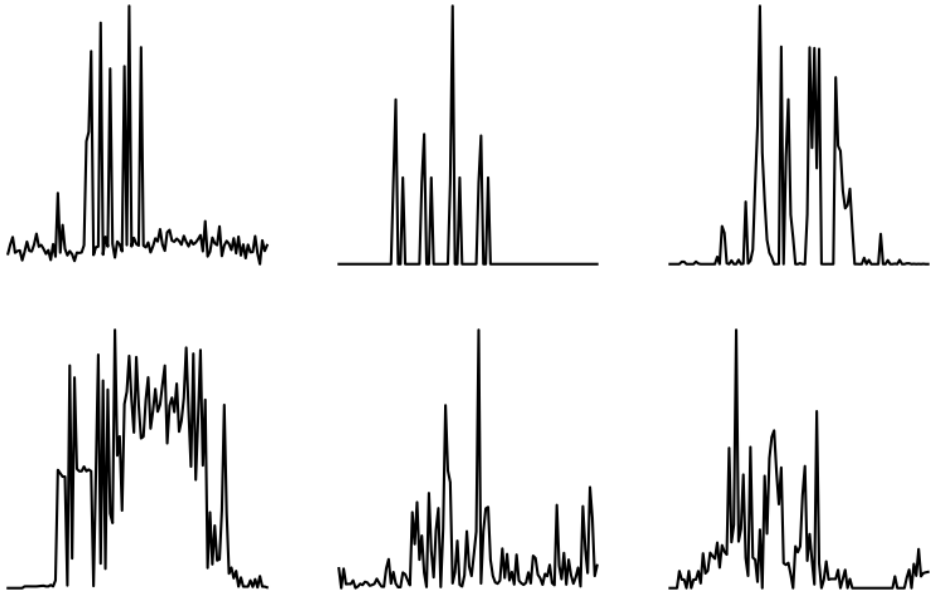


Figure 5: 6 companies (out of 13) classified and *Cass 1* (suspicious) against the chosen governments transitions.

ultimately leading to a more robust and reliable system for time series classification in practical applications.

5 Acknowledgements

We acknowledge support of the Slovenian Research Agency for financing this work through projects MR-Joksimović (53925) and research core funding for the programmes Complex Networks (No. P1-0383) and Knowledge Technologies (No. P2-0103). We also thank KPK for sharing the data and their challenges with us.

References

[1] Joksimović, J.: Patterns discovery in Slovenian public spending: a data-driven approach to corruption detection. Ph.D. thesis (11 2023). https://doi.org/10.13140/RG.2.2.35480.81926

[2] Joksimović, J., Perc, M., Levnajić, Z.: Self-organization in slovenian public spending. Royal Society Open Science **10**(8) (2023). https://doi.org/https://doi.org/10.1098/rsos.221279

[3] Kaufman, L., Rousseeuw, P.: Finding Groups in Data: An Introduction To Cluster Analysis (01 1990). https://doi.org/10.2307/2532178

[4] Levnajić, Z., Ženko, B., Joksimović, J.: Detecting corruption in slovenian public spending from temporal data (11 2023)

[5] Lzp, J.: Distance measures and linkage methods in hierarchical clustering (2019)

[6] MacQueen, J.: Some methods for classification and analysis of multivariate observations (1967)

[7] Wang, X., Smith, K., Hyndman, R.: Characteristic-Based Clustering for Time Series Data. Data Mining and Knowledge Discovery **13**(3), 335–364 (2006). https://doi.org/10.1007/s10618-005-0039-x, https://doi.org/10.1007/s10618-005-0039-x

Towards a randomized algorithm for the maximum stable set problem

Omkar Bihani[1], Janez Povh [1,2], and Janez Žerovnik[1,2*]

[1] Rudolfovo - Science and Technology Centre Novo Mesto, Podbreznik 15, 8000 Novo mesto, Slovenia.

[2] University of Ljubljana, Faculty of Mechanical Engineering, Aškerčeva cesta 6, 1000 Ljubljana, Slovenia;

omkar.bihani@rudolfovo.eu, janez.povh@rudolfovo.eu, janez.zerovnik@fs.uni-lj.si

Abstract. *Nearly fifty years ago, Petford and Welsh introduced a sequential heuristic algorithm for (approximately) solving the NP-hard graph coloring problem. The algorithm is based on the antivoter model and mimics the behaviour of a physical process based on a multi-particle system of statistical mechanics. It was later shown that the algorithm can be implemented in massively parallel model of computation. With some natural modifications, the algorithm was successfully applied to various problems similar to the graph coloring including the frequency assignment problems and the graph clustering problem. In this paper, we report experimental results that demonstrate this algorithm as a competitive heuristics for computing the maximum stable set in a graph.*

Keywords. stable set problem, randomized heuristics, graph coloring

1 Introduction

The maximum stable set problem (MSS), or, the maximum independent set, is a well-known combinatorial optimization problem. For a given undirected graph $G = (V, E)$, the MSS asks what is the cardinality of the largest subset of vertices such that the graph induced on this subset has no edge. This value is called the stability number of the graph and is usually denoted by $\alpha(G)$. Computation of $\alpha(G)$ is an NP-complete problem contained in the famous Karp's list from 1972.

Therefore, unless the set of P and NP problems are the same, we can not solve the MSS problem to optimality in computing time that is polynomial function of the length of the encoding of the instance. However, we can see several real-life applications, for instance, in coding theory [7], in bioinformatics [5], scheduling [21], etc.

Therefore, there is much interest in heuristic algorithms which can find near-optimal solutions within reasonable running times for a large proportion of instances.

In a review of algorithms for the maximum clique problem (which is equivalent to MSS, computed on the graph complement) [23], the authors compared 21 different heuristics. They showed that the general swap-based multiple neighborhood tabu search approach from [24] yielded the best results. More recent review [14] additionally points out that the local search based algorithm from [13] is also extremely effective. We also recommend the reader to read recent paper for more detailed state of the art in solving the MSS exactly and approximately [12].

*Corresponding author.

Another closely related problem is the graph coloring problem. More precisely, a proper coloring of (vertices) of a graph is a partition of the vertex set in which each set is stable. On the other hand, given any stable set, there is a proper coloring in which all the vertices are assigned the same color. Hence, any coloring algorithm outputs stable sets, however they need not be close to maximum. Usually, the coloring algorithms tend to provide colors sets of approximately the same size. However, if the coloring algorithm tends to favor one color, it will produce uneven color sets and the largest may be a good approximation of the maximum independent set. In the work reported here we investigate how this idea can be used a particular heuristics for graph coloring.

Nearly fifty years ago, Petford and Welsh proposed a randomized algorithm for graph coloring, based on the antivoter model of Donnely and Welsh [3]. Experiments on a model of random 3-colorable graphs shown that there are some combinations of parameters of random graphs that are extremely hard for their algorithm, while otherwise their algorithm was running on average in linear time. Having no theoretical explanation, Petford and Welsh write that the curious behavior "is not unlike the phenomenon of phase-transition which occurs in the Ising model, Potts model and other models of statistical mechanics" [16]. With some natural modifications, the heuristics was successfully applied to the k-coloring problem [28, 19] and to the problem of computing the chromatic number of a graph [25]. With some straightforward modifications, the algorithm was also very competitive on frequency assignment problems [4, 20, 29]. We have recently designed an algorithm for clustering that is motivated by this coloring algorithm. The results were very promising [9].

An interesting feature of the heuristics is that is equivalent to the operation of the Boltzman machine at fixed temperature [18]. The experimentally shown optimal speedup in a massively parallel model of computation [26, 27], together with the basic motivation in models of statistical physics, gives rise to the idea that in may be relevant and in some form usable in quantum models of computing [8].

Threshold phenomena have attracted a lot of attention in the context of random combinatorial problems [6] and in theoretical physics [17]. In statistical physics, phase transitions have been studied for more than a century. Let us only mention here that the spin glasses, a purely theoretical concept, has triggered a new branch of theoretical physics that resulted in a Nobel prize attributed to Giorgio Parisi in 2022 [15]. Graph coloring with $k = 3$ colors has been considered in several papers, see [2, 1] and the references there. In contrast to graph classes of k colorable graphs used in [16, 8], the usual random graph model considered are graphs $G(n, p)$ where each of the possible $\frac{n(n-1)}{2}$ edges appears independently with probability p. Not surprisingly, for 3-coloring, it is found that the critical mean degree where the phase transition occurs is around $\alpha_{crit} \approx 4.7$, for example the estimate 4.703 was put forward in [1]. In the same paper, the analysis implies that the hardest instances are among graphs with average degree between 4.42 and α_{crit}. Of course, if we are going to apply the coloring algorithm or its modification to the stable set problem, we must be aware of the existence of the critical regions and thus expect that the algorithm will perform badly at some combinations of parameters.

In this paper we report on testing some ideas on how to modify the coloring algorithm of Petford and Welsh to provide hopefully good approximations for the maximum stable set problem. The first results are promising, as we obtain optimal solutions for some instances. However, we observe that several parameters may in some cases substantially change the performances, indicating the important issues that the design of a future randomized heuristics should address.

The rest of paper is organized as follows. In Section 2 we provide some basic informa-

tion on the algorithm and on the graph classes used in the experiment. the new, modified algorithm is outlined in Section 3 In section 4 we present the experiments and comment the preliminary results. In the last section we add the discussion on the inherent paralelism of the randomized heuristics that may lead to a massively parallel implementation in a suitable model of computation.

2 The problems and the basic algorithm

The basic objects in graph theory are graphs, defined as a tuple $G = (V(G), E(G))$, where $V = V(G)$ is a set of vertices (nodes), and $E(G)$ is a set of edges. Edges may be defined as pairs of vertices the edge connects. We recall some definitions below, for other we refer to [22]. We say a mapping $c : V(G) \to \mathcal{N}$ is a *proper coloring* of G if it assigns different colors to adjacent vertices. Any mapping $c : V(G) \to \mathcal{N}$ will be called a *coloring*, and will be considered as one of the feasible solutions of the problem.

We need to use "the maximum stable set", not the maximal, since graph can have many maximal stable sets with various sizes. The MSS asks for cardinality, not necessarily for the set itself.

The maximum stable set problem.

PROBLEM : **The maximum stable set problem.**
Input: graph G
Task: find the cardinality of the maximum stable set in G.

Graph coloring problems. Coloring of the vertices of a graph is usually asked either in the form of optimization or in the form of decision problem. The first asks for the chromatic number, i.e. the minimal number of colors that allows a proper coloring.

PROBLEM : **Graph coloring optimization problem).**
Input: graph G,
Task: find the chromatic problem of G.

The k-coloring decision problem is a well known NP-complete problem for $k \geq 3$. It reads as follows:

PROBLEM : **Graph coloring (decision problem).**
Input: graph G, integer k
Question: is there a proper k-coloring of G?

The algorithm of Petford and Welsh for k-coloring decision problem.

Let us first introduce the cost function $E(c)$ to be the number of *bad* edges, i. e. edges with both ends colored by the same color by coloring c. Proper colorings are exactly the colorings for which $E(c) = 0$ and finding a coloring c with $E(c) = 0$ is equivalent to answering the above decision problem where the coloring constructed is a *witness* c proving the correctness of the answer.

The basic algorithm [16] starts with a random initial 3 coloring of the input graph and then applies an iterative process. In each iteration a vertex creating a conflict is chosen at random. The random distribution is uniform among the bad vertices and the chosen

vertex is recolored according to some probability distribution. The color distribution favors colors which are less represented in the neighborhood of the chosen vertex, see the expression (1) below. The algorithm has a straightforward generalization to k coloring (taking $k = 3$ gives the original algorithm) [28].

In a pseudo language the algorithm of Petford and Welsh can be written as

Algorithm 1 Petford-Welsh (PW) algorithm

```
color vertices randomly with colors 1, 2, ..., k
while not stopping condition do
    select a bad vertex v (randomly)
    assign a new color to v
end while
```

Bad vertex is selected uniformly random among vertices which are endpoints of some bad (e.g. monochromatic) edge. A new color is assigned at random. The new color is taken from the set $\{1, 2, \ldots, k\}$. Sampling is done according to probability distribution defined as follows:

The probability p_i of color i to be chosen as a new color of vertex v is proportional to

$$p_i \approx \exp(-S_i/T) = b^{-S_i}, \tag{1}$$

where S_i is the number of edges with one endpoint at v and with color i assigned to the other endpoint. Petford and Welsh used 4^{-S_i} which is equivalent to using $T \simeq 0.72$ in (1). (Because $\exp(-x/T) = 4^{-x}$ implies $T \simeq 0.72$.)

The stopping criteria is either reaching a time limit (in terms of the number of calls to the function which computes a new color) or if a proper coloring is found. If no proper coloring is found, the solution with minimal cost $E(c)$ is reported and can be regarded as an approximate solution to the problem. However, there is no guarantee on the quality of the solution known.

As already explained, the original algorithm of Petford and Welsh uses probabilities proportional to 4^{-S_i}, which corresponds to $T \approx 0.72$. Larger values of T result in higher probability of accepting a move which increases the number of bad edges. Clearly, a very high T results in chaotic behavior similar to a pure random walk among the colorings ignoring the their energy. On the other hand, with low values of T, the algorithm behaves very much like iterative improvement, quickly converging to a local minimum. Therefore, the parameter T may called *temperature*. Further arguments for this wording are based on analogies to the Boltzmann machine and simulated annealing algorithm [11, 30].

3 The modified algorithm for stable set

As already discussed, as often is the case, the coloring algorithms tend not to favor any particular color. This is obviously true for the Petford Welsh algorithm, in which the probability of choosing a color depends only on S_i, the number of neighbors of the vertex currently colored by color i. Because in the maximum stable set problem, we need just one large stable set, we modify the probability distribution in a way that will favor one color, without loss of generality this will be color 1.

In search for a suitable probability distribution, we have checked some versions and found that the formula (2) below is performing well, at least as good as some other ideas

we had. (It remains to explore some other alternatives.) In the modified algorithm, we change the definition of acceptance probabilities (1) to

$$p_i \approx b^{F_i} b^{-S_i}, \tag{2}$$

where F_i is the bias factor, defined as

$$F_i = BIAS \begin{cases} LARGE & \text{if } i = 1, \\ 1 & \text{otherwise} \end{cases}$$

The value of $LARGE$ reflects the preference for the first color over the others. Say, for a graph colored with three colors, the bias vector $\vec{F} = [LARGE, 1, 1]$, with $LARGE$ determined by the relative preference for the first color.

The value of $LARGE$ is determined by the equation below:

$$LARGE = 1 - \frac{\log c}{\log b} \tag{3}$$

where $c = \frac{1-p}{(k-1)p}$, p is the probability of choosing the first color and $b =$ base.

Thus we have changed the acceptance probabilities to favorize color 1. As expected, the experiments confirm that the size of the set of vertices colored by 1 tends to be the largest, and that it gives in some cases a good approximation for the maximum stable set.

The modification just explained inherits a major problem because the algorithm of Petford and Welsh uses the information on the number of colors, k. In the stable set problem, the chromatic number of the graph is of course not given, and any algorithm must work without this information. We have observed that indeed the modified algorithm's performance heavily depends on k. For more details, see the experiments below. Our first and simple solution was to run a batch of experiments to obtain an approximation of the chromatic number as a preprocessing, and then run the modified algorithm for several k close to the estimated chromatic number. More advanced heuristics may possibly be developed in a way, similar to the adaptation of Petford Welsh algorithm that computed approximation of the chromatic number [25].

Another well-known and therefore expected phenomena is the existence of critical regions which in practice means that the performance of the heuristics may in some cases heavily depend on the temperature T, or, in other words, on the choice of the exponent basis. We start with base 4, as in the original work [16], and start investigation on how this parameter is important for the performance of our algorithm.

4 Experiments

In the preliminary experiments we have tested the ideas on randomly generated instances. Due to space limitations we omit the results here, and consider the instances that have known maximum stable sets (given as "True card" in Table 1). The instances for the experiments are chosen from the second DIMACS Implementation Challenge [10].

4.1 Preprocessing to find a suitable number of colors

As the Algorithm 1 and its modification need the number of colors k as a given parameter, we run several experiments with various k. (The chromatic number of the instances is not always known.) The observation was that the algorithms provide better solutions when

run at k that is close to minimal number of colors with which Algorithm 1 was able to find a proper coloring. Therefore, we introduced a simple preprocessing phase that provides an approximation to the chromatic number, described in Algorithm 2. The maximum number of iterations to get the $best\_k$ was set to $2000n$ where n is the number of vertices.

Algorithm 2 Estimation of k by binary search

Input: Graph g, initial upper bound UB = degree of maximum node + 1, steps
Initialize: $LB = 0$, $best_k = UB$
while $LB \leq UB$ **do**
    $k \leftarrow \lfloor \frac{LB+UB}{2} \rfloor$
    PWG $\leftarrow$ PetfordWelshColoring(g, k, steps)
    if good_coloring(PWG) **then**
        $best_k \leftarrow k$
        $UB \leftarrow k - 1$
    else
        $LB \leftarrow k + 1$
    end if
end while
Return: $best_k$

4.2 Changing the acceptance probabilities to favourize color 1

It is obvious that a proper coloring provided by Algorithm 1 gives k stable sets. As the colors are chosen with probabilities independent of the name of the color, it is natural to expect that the colors sets will tend to be of approximately equal size. As we want to have at least one large color set, it is natural to introduce a mechanism that will favourize one color, as we did by formula (2) when introducing the modified algorithm. Interesting observation is that the modification did not always yield a better result than the Algorithm 1, see Table 1.

Among many possible natural choices, we decided to consider three cases where the probability of coloring the bad vertex with the first color is 0.5, 0.7, and 0.9, while the other colors are selected with equal probability. The value of $LARGE$ is set following equation (3) for these three cases. To get the cardinalitites, we set the max. iters. as $200n$ for n nodes in the graph.

Table 1 shows the best cardinalities found by PW when colored with $best\_k$ colors found by algorithm 2 at respective probabilities for base = 4 and $BIAS = 1.0$

Table 1: Best cardinalities found at different probabilities. base = 4, $BIAS = 1.0$

name	N	Edges	deg	best_k	PW_card	mPW_card (0.5)	mPW_card (0.7)	mPW_card (0.9)	True card.
brock200_1	200	5066	50.7	17	17	16	17	17	21
brock200_2	200	10024	100.2	34	9	9	9	3	12
brock200_3	200	7852	78.5	25	12	10	11	2	15
brock200_4	200	6811	68.1	23	13	13	13	13	17
brock800_1	800	112095	280.2	79	15	16	16	15	23
torus11	121	242	4.0	3	46	46	44	46	55
torus13	169	338	4.0	3	63	63	64	65	78
torus15	225	450	4.0	3	82	86	91	85	105
torus17	289	578	4.0	3	106	105	108	111	136
torus19	361	722	4.0	3	134	133	135	137	171
san200_0_7_1	200	5970	59.7	14	16	16	15	15	30
san200_0_7_2	200	5970	59.7	22	13	13	13	13	18
sanr200_0_7	200	6032	60.3	20	14	14	14	14	18

We next run a batch of experiments where we use k ($k+1, k-1, k-2$) computing by Algorithm 2. We also vary $BIAS$ from the set $[0.0, 0.2, 0.4, 0.6, 0.8, 1.0]$. Summary of the results is given in Table 2. A list of runs that gave the best result obtained is given with the set of parameters. Observe that several cases, the optimal values have been found.

Table 2: Combinations of parameters at the best cardinality found at base = 4.
The numbers under mPW indicate the probability of selecting the first color.

name	best_k	mPW (0.5)	comb (0.5)	mPW (0.7)	comb (0.7)	PW (0.9)	comb (0.9)	True card
brock200_1	17	18	('17', '0.4', False, 18, 0, 21869) ('17', '0.8', False, 18, 0, 27247) -	18	('15', '0.2', False, 18, 0, 39692) ('16', '0.0', False, 18, 0, 32876) ('16', '0.4', False, 18, 0, 33758)	18	('16', '0.2', False, 18, 0, 32912) ('17', '0.4', False, 18, 0, 22511) -	21
brock200_2	34	**12**	('35', '0.6', False, 12, 0, 21257) - -	**12**	('33', '0.2', False, 12, 0, 25330) ('33', '0.6', False, 12, 0, 34081) ('35', '1.0', False, 12, 0, 19790)	**12**	('33', '0.4', False, 12, 0, 34721) ('33', '0.6', False, 12, 0, 28239) ('33', '0.8', False, 12, 0, 27284)	12
brock200_3	25	**15**	('23', '0.6', False, 15, 0, 33578) ('24', '0.0', False, 15, 0, 30801) ('24', '0.4', False, 15, 0, 35923)	**15**	('23', '0.0', False, 15, 0, 39040) ('23', '0.2', False, 15, 0, 39155) ('23', '0.4', False, 15, 0, 37383)	**15**	('23', '0.4', False, 15, 0, 36373) ('24', '0.0', False, 15, 0, 36104) ('24', '0.2', False, 15, 0, 34406)	15
brock200_4	23	14	('22', '0.2', False, 14, 0, 34852) ('22', '0.6', False, 14, 0, 29332) ('22', '0.8', True, 14, 0, 28674)	14	('22', '0.6', True, 14, 0, 16087) ('22', '0.8', True, 14, 0, 21156) ('24', '0.6', True, 14, 0, 8309)	15	('21', '0.4', False, 15, 0, 37431) - -	17
brock800_1	79	17	('77', '0.6', False, 17, 0, 159936) ('79', '0.2', True, 17, 0, 58178) ('79', '0.4', True, 17, 0, 63250)	17	('77', '0.4', False, 17, 0, 159936) ('79', '0.0', False, 17, 0, 156845) -	17	('77', '0.2', False, 17, 0, 159936) ('77', '0.4', False, 17, 0, 159936) -	23
torus11	3	47	('3', '0.4', False, 47, 0, 2858) ('3', '0.6', False, 47, 0, 2405) -	47	('3', '0.0', True, 47, 0, 1460) ('3', '0.6', False, 47, 0, 2408) ('3', '0.8', True, 47, 0, 2586)	48	('3', '0.2', True, 48, 0, 3238) ('3', '0.6', True, 48, 0, 2759) -	55
torus13	3	64	('2', '0.2', False, 85, 21, 31009) ('3', '0.4', False, 64, 0, 5301) ('3', '0.6', True, 64, 0, 1530)	67	('3', '0.8', False, 67, 0, 3278) - -	66	('3', '0.6', False, 66, 0, 3910) - -	78
torus15	3	86	('2', '0.0', False, 113, 27, 42410) ('2', '0.4', False, 113, 27, 42398) ('3', '1.0', False, 86, 0, 5966)	91	('3', '1.0', False, 91, 0, 7832) - -	87	('2', '0.0', False, 114, 27, 42993) - -	105
torus17	3	115	('2', '0.4', False, 145, 30, 57197) ('2', '0.6', False, 145, 30, 57453)	117	('2', '0.0', False, 145, 28, 56511) ('2', '0.4', False, 145, 28, 55930)	118	('2', '0.0', False, 145, 27, 56932) -	136
torus19	3	149	('2', '0.0', False, 182, 33, 72047) - -	148	('2', '0.0', False, 182, 34, 71998) ('2', '0.2', False, 182, 34, 72015) ('2', '0.6', False, 181, 33, 71624)	147	('2', '0.6', False, 183, 36, 71775) - -	171
san200_0_7_1	14	**30**	('12', '0.6', False, 30, 0, 11307) -	**30**	('12', '0.0', False, 30, 0, 12680) ('12', '0.4', False, 30, 0, 15736)	**30**	('12', '0.0', False, 30, 0, 12216) -	30
san200_0_7_2	22	14	('21', '0.0', False, 14, 0, 12144) ('21', '0.4', False, 14, 0, 10613) ('21', '0.8', False, 14, 0, 12519)	14	('21', '0.2', False, 14, 0, 14403) ('21', '0.4', False, 14, 0, 12277) ('21', '0.6', False, 14, 0, 13010)	14	('21', '0.8', False, 14, 0, 13356) ('23', '0.2', True, 14, 0, 2810) -	18
sanr200_0_7	20	16	('18', '0.0', False, 16, 0, 38820) ('20', '0.2', False, 16, 0, 21313) -	16	('19', '0.4', False, 16, 0, 31910) - -	15	('20', '0.0', False, 15, 0, 26294) ('20', '0.4', False, 15, 0, 23711) ('20', '0.8', True, 15, 0, 9603)	18

In Table 2, the combinations is a tuple describing ('k','$BIAS$','good color','# vertices of dominant color','# bad vertices of dominant color' '# iterations')

4.3 Further parameter adjustments, base (temperature)

As already mentioned, it is well known that the performance Algorithm 1 may heavily depend on the base (temperature) chosen. Thus we perform a batch of experiment with varying value of base. First, we get the best_k across different bases by using Algorithm 2 as presented in Table 3. Next, we apply the modified algorithm on the values from table 3 to find the stable set and calculate the cardinality at various bases. The results for base = 8 and $BIAS = 1.0$ is shown in Table 4, where we can see the differences in the values of cardinality with those in Table 1 where base = 4 for the same $BIAS$ value. Finally, Table 5 highlights the best cardinality achieved at different bases. Here, due to space constraints, the specific combinations that yielded these cardinalities are not provided.

5 Conclusions

We wish to note that the randomized algorithm we put forward is of particular interest because of its high inherent parallelism. In [26], a massively parallel version of Algorithm

Table 3: Best k and # iterations found by PW at different bases.

name	N	Edges	deg	3	4	5	6	7	8	9	10
brock200_1	200	5066	50.7	k = 18 13597	k = 17 47327	k = 17 20471	k = 16 25336	k = 15 36667	k = 15 60763	k = 15 10806	k = 15 16977
brock200_2	200	10024	100.2	k = 39 74447	k = 34 83006	k = 32 29964	k = 30 195013	k = 30 198093	k = 29 77822	k = 29 21326	k = 28 86223
brock200_3	200	7852	78.5	k = 28 257910	k = 25 240388	k = 24 53569	k = 22 73560	k = 22 137067	k = 21 61748	k = 21 106910	k = 22 17325
brock200_4	200	6811	68.1	k = 25 63289	k = 23 46393	k = 22 11475	k = 21 28599	k = 20 57640	k = 20 18581	k = 20 14127	k = 19 45250
brock800_1	800	112095	280.2	k = 93 854308	k = 79 149280	k = 72 432807	k = 68 119385	k = 66 214438	k = 65 245306	k = 63 667428	k = 62 606459
torus11	121	242	4.0	k = 3 1635	k = 3 3166	k = 3 1668	k = 3 1590	k = 3 4999	k = 3 1428	k = 3 982	k = 3 4755
torus13	169	338	4.0	k = 3 3260	k = 3 6919	k = 3 8317	k = 3 1379	k = 3 3708	k = 3 1476	k = 3 1347	k = 3 2219
torus15	225	450	4.0	k = 3 3599	k = 3 3062	k = 3 2822	k = 3 1189	k = 3 5642	k = 3 11005	k = 3 26029	k = 3 1183
torus17	289	578	4.0	k = 3 7712	k = 3 2818	k = 3 15008	k = 3 1883	k = 3 7544	k = 3 17151	k = 3 9480	k = 3 26335
torus19	361	722	4.0	k = 3 5656	k = 3 25305	k = 3 5243	k = 3 13268	k = 4 283	k = 3 22018	k = 3 5985	k = 3 22077
san200_0_7_1	200	5970	59.7	k = 14 3859	k = 14 4655	k = 14 3736	k = 14 6390	k = 15 2026	k = 14 3139	k = 14 2738	k = 14 11847
san200_0_7_2	200	5970	59.7	k = 22 3240	k = 22 3743	k = 22 5216	k = 22 3532	k = 22 8782	k = 22 2539	k = 21 29505	k = 21 18371
sanr200_0_7	200	6032	60.3	k = 22 72442	k = 20 67893	k = 19 161607	k = 18 34720	k = 18 14261	k = 18 9734	k = 17 48078	k = 17 120394

Table 4: Best cardinalities found at different probabilities. base = 8, $BIAS = 1.0$

name	N	Edges	deg	best_k	PW_card	mPW (0.5)	mPW (0.7)	mPW (0.9)	True card
brock200_1	200	5066	50.7	15	17	17	18	17	21
brock200_2	200	10024	100.2	29	11	9	**12**	9	12
brock200_3	200	7852	78.5	21	13	14	12	**15**	15
brock200_4	200	6811	68.1	20	14	13	14	13	17
brock800_1	800	112095	280.2	65	19	18	17	19	23
torus11	121	242	4.0	3	47	46	45	45	55
torus13	169	338	4.0	3	66	63	64	61	78
torus15	225	450	4.0	3	81	86	90	84	105
torus17	289	578	4.0	3	103	108	109	107	136
torus19	361	722	4.0	3	130	135	136	137	171
san200_0_7_1	200	5970	59.7	14	22	22	16	16	30
san200_0_7_2	200	5970	59.7	22	13	13	14	13	18
sanr200_0_7	200	6032	60.3	18	15	15	15	15	18

Table 5: Best cardinalities found at each base for various graphs

name	N	Edges	deg	3	4	5	6	7	8	9	10	True card.
brock200_1	200	5066	50.7	18	18	18	20	19	20	20	20	21
brock200_2	200	10024	100.2	11	**12**	**12**	**12**	**12**	**12**	**12**	**12**	12
brock200_3	200	7852	78.5	**15**	**15**	**15**	**15**	**15**	**15**	**15**	**15**	15
brock200_4	200	6811	68.1	14	15	**17**	16	**17**	**17**	16	15	17
brock800_1	800	112095	280.2	16	17	19	19	20	20	20	19	23
torus11	121	242	4.0	49	48	48	48	49	48	49	49	55
torus13	169	338	4.0	69	67	70	69	68	68	68	67	78
torus15	225	450	4.0	92	91	90	90	90	92	91	91	105
torus17	289	578	4.0	114	118	119	119	120	120	120	120	136
torus19	361	722	4.0	145	149	152	153	153	154	153	154	171
san200_0_7_1	200	5970	59.7	**30**	**30**	**30**	**30**	29	**30**	**30**	**30**	30
san200_0_7_2	200	5970	59.7	14	14	17	14	15	15	17	15	18
sanr200_0_7	200	6032	60.3	15	16	17	17	17	17	17	17	18

1 was proposed. The naive algorithm used in [26] was later improved in [27] by a version that runs in two phases thus avoiding the looping that may appear at some configurations within the instance. In short, simulations have shown that the speedup of the coloring algorithm was best possible, namely the linear time complexity of sequential version was sped up to constant time complexity in a model with n processors. We believe that the modified algorithm can enjoy the same speed up.

6 Acknowledgments

The research of the authors was partially funded by the Slovenian Research and Innovation Agency (ARIS) through the annual work program of Rudolfovo, and by the Slovenian Ministry of Higher Education, Science and Innovation, ARIS, and by the European Union -NextGenerationEU through the DIGITOP project. The third author was also partly supported by ARIS through the research grants P2-0248 and J1-4031.

References

[1] Boettcher, S.; Percus, A.G. Extremal optimization at the phase transition of the three-coloring problem. Phys. Rev. E, volume 69, page 066703, 2004.

[2] Culberson, J.; Gent, I. Frozen development in graph coloring. Theoretical Computer Science, volume 265, pages 227–264, 2001.

[3] Donnelly, P.; Welsh, D. The antivoter problem: random 2-colourings of graphs. In Graph Theory and Combinatorics, pages 133–144, Academic Press, London, 1984.

[4] Chamaret, B.; Ubeda, S.; Žerovnik, J. A randomized algorithm for graph colouring applied to channel allocation in mobile telephone networks. In Proceedings of the 6th International Conference on Operational Research KOI'96, pages 25–30, Croatian Operational Research Society Zagreb, 1996.

[5] Malod-Dognin, N.; Andonov, R. Maximum Cliques in Protein Structure Comparison. In Experimental Algorithms, pages 106–117, Springer Berlin Heidelberg, 2010.

[6] Dubois, O.; Monasson, R.; Selman, B.; Zecchina, R. Editorial. Theoretical Computer Science, volume 265, page 1, 2001.

[7] Etzion, T.; Ostergard, P. Greedy and heuristic algorithms for codes and colorings. IEEE Transactions on Information Theory, volume 44, pages 382–388, 1998.

[8] Gabrovšek, B.; Žerovnik, J. A fresh look to a randomized massively parallel graph coloring algorithm. Croatian Operational Research Review, volume 15, pages 105–117, 2024.

[9] Ikica, B.; Gabrovšek, B.; Povh, J.; Žerovnik, J. Clustering as a dual problem to colouring. Computational and Applied Mathematics, volume 41, page 147, 2022.

[10] Johnson, D.J.; Trick, M.A. Cliques, Coloring, and Satisfiability: Second DIMACS Implementation Challenge, October 11-13, 1993. American Mathematical Society, 1996.

[11] Kirckpatrick, S.; Gellat, C.D. Jr.; Vecchi, M.P. Optimization by simulated annealing. Science, volume 220, pages 671–680, 1983.

[12] Krpan, A.; Povh, J.; Pucher, D. Quantum computing and the stable set problem. arXiv preprint arXiv:2405.12845, 2024.

[13] Lamm, S.; Sanders, P.; Schulz, C.; Strash, D.; Werneck, R.F. Finding near-optimal independent sets at scale. Journal of Heuristics, volume 23, pages 207–229, 2015.

[14] Marino, R.; Buffoni, L.; Zavalnij, B. A short review on novel approaches for maximum clique problem: From classical algorithms to graph neural networks and quantum algorithms. arXiv preprint arXiv:2403.09742, 2024.

[15] Mezard, M. Spin glasses and optimization in complex systems. Europhysics News, volume 53, pages 15–17, 2022.

[16] Petford, A.; Welsh, D. A randomized 3-coloring algorithm. Discrete Mathematics, volume 74, pages 253–261, 1989.

[17] Philathong, H.; Akshay, V.; Samburskaya, K.; Biamonte, J. Computational phase transitions: benchmarking Ising machines and quantum optimisers. Journal of Physics: Complexity, volume 2, page 011002, 2021.

[18] Shawe-Taylor, J.; Žerovnik, J. Boltzmann Machines with Finite Alphabet. In Artificial Neural Networks 2, volume 1, pages 391–394, Brighton, U.K., 1992.

[19] Shawe-Taylor, J.; Žerovnik, J. Analysis of the Mean Field Annealing Algorithm for Graph Colouring. Journal of Artificial Neural Networks, volume 2, pages 329–340, 1995.

[20] Ubeda, S.; Žerovnik, J. A randomized algorithm for a channel assignment problem. Speedup, volume 11, pages 14–19, 1997.

[21] Weide, O.; Ryan, D.; Ehrgott, M. An iterative approach to robust and integrated aircraft routing and crew scheduling. Computers & Operations Research, volume 37, pages 833–844, 2010.

[22] Wilson, R.; Watkins, J. Graphs: An Introductory Approach. John Wiley & Sons, Inc., New York, 1990.

[23] Wu, Q.; Hao, J-K. A review on algorithms for maximum clique problems. European Journal of Operational Research, volume 242, pages 693–709, 2015.

[24] Yan, J.; Hao, J-K. General swap-based multiple neighborhood tabu search for the maximum independent set problem. Engineering Applications of Artificial Intelligence, volume 37, pages 20–33, 2015.

[25] Žerovnik, J. A randomized heuristical algorithm for estimating the chromatic number of a graph. Information Processing Letters, volume 33, pages 213–219, 1989.

[26] Žerovnik, J. A parallel variant of a heuristical algorithm for graph coloring. Parallel Computing, volume 13, pages 95–100, 1990.

[27] Žerovnik, J.; Kaufman, M. A parallel variant of a heuristical algorithm for graph coloring - corrigendum. Parallel Computing, volume 18, pages 897–900, 1992.

[28] Žerovnik, J. A randomized algorithm for k-colorability. Discrete Mathematics, volume 131, pages 379–393, 1994.

[29] Žerovnik, J. On the convergence of a randomized algorithm frequency assignment problem. Central European Journal for Operations Research and Economics, volume 6, pages 135–151, 1998.

[30] Žerovnik, J. On temperature schedules for generalized Boltzmann machine. Neural Networks World, volume 3, pages 495–503, 2000.

Classification of bug severity with lexicon approaches

Fatima Aziz
Jožef Stefan Institute, Jožef Stefan International Postgraduate School
Jamova cesta 39, 1000 Ljubljana, Slovenia
fatima.aziz6655@gmail.com
Martin Žnidaršič
Jožef Stefan Institute, Jožef Stefan International Postgraduate School
Jamova cesta 39, 1000 Ljubljana, Slovenia
martin.znidarsic@ijs.si

Abstract. *Bug severity classification is one of important and time consuming parts of software bug resolution processes. This task is usually automated with supervised machine learning methods, but results of our preliminary investigation about suitability of lexicon-based approaches in this domain indicate that development of a specialized lexicon for this purpose would also make sense. In this paper we present the development of such a lexicon and its empirical evaluation alongside similar approaches and already publicly available lexicons from a related domain of sentiment analysis. Experimental results of classification performance are very dataset dependant and indicate also usefulness of the sentiment analysis lexicons.*
Keywords. lexicon-based classification, software bug severity, software bug report, software maintenance, sentiment analysis

1 Introduction

Bugs in software systems refer to faults that prevent the systems to operate as intended. Identification and resolution of bugs is thus crucial to ensure proper functionality and quality of software. In this process, efficient bug triaging, usually according to perceived bug severity is a very important aspect. Bug severity is assessed based on bug reports, in which textual descriptions of bugs are one of the most important components. Automation of such text-based bug severity classification is mostly approached with machine-learning methods [7], [3], either with classic methods, or with methods based on deep neural networks and large language models, which are recently an ever more frequent choice.

A simpler commonly used baseline in many problem domains of text-based classification is the lexicon-based approach, which makes use of predefined lexicons of terms that are indicative of a specific target class. Besides serving as a simple and computationally non-demanding baseline, lexicon based approaches are sometimes useful also due to their so-called *white box* nature, meaning that the reasons for their results can always be tracked-back and fully understood. Interestingly, there are no publicly available lexicons for bug severity classification, although there was some research done on this topic [2].

Our preliminary investigation [1] implies that development of such a lexicon could be useful. In this paper, we report on the process of its development and present a quantitative and qualitative comparison of its performance in relation to similar existing approaches and resources. The lexicons and lexicon creation methods that we take into account are presented in Section 2, the experiments are described in Section 3, results are shown in section 4, conclusions is provided in Section 5.

2 Lexicon-based approaches

2.1 Lexicons

In this section we briefly describe the four lexicons that were used in our experiments: two lexicon creation approaches that we followed to create bug severity lexicons and two public lexicons for a different, but somewhat related domain of sentiment analysis.

2.1.1 A threshold-based approach

Here we provide a brief description of our previously proposed threshold based lexicon development approach. For details, see the original publication [1]. The approach starts by transforming the training dataset by filtering out punctuation and stop words and splitting into individual words. Next, we count word occurrences in the Summary column for both Severe and Non-severe categories, generating two word lists that reflected word frequencies in each category. We calculated the ratio of individual word appearances in both categories and defined thresholds for severe and non-severe classifications. By exploring various combinations of these thresholds, we generated subsets of the main word list to identify the most significant words. The thresholds yielding the highest F1 scores on a separate validation set are then selected to create the output lexicon.

2.1.2 Linear SVM-based approach

We also created a lexicon based on a linear SVM classifier following the approach from a paper by Islam et al. [6]. The summary text column from the training dataset was pre-processed with NLP techniques to obtain a corpus of words. The CountVectorizer method was used to transform the processed text into a numerical word matrix. Through training and testing with different values of parameter c on a validation set, the best parameter value was determined in terms of the F1 score. The linear SVM model with the best performing c parameter was was used to gather the information on the coefficients, with positive coefficients used as indicative of the target class (severe), and negative coefficients for the opposite class (non-severe). Consequently, we established two distinct word lists: for severe category comprising words with positive coefficients, and for non-severe category, including words with negative coefficients. Static lexicon was generated for both categories and used in the lexicon-based classifier.

2.1.3 Bing Liu's Opinion Lexicon

A study was conducted utilizing the publicly available Bing Liu sentiment lexicon [4] to classify software bug severity. The Bing Liu lexicon consists of two text files: one containing positive words and the other containing negative words. The positive file includes

words with positive sentiments, while the negative file contains words with negative sentiments. These two files were used in place of our created lexicons, and we subsequently applied our classification method, which is based on the intersection of word counts, to classify the bug reports. In general, Bing Liu's positive lexicon comprises words and phrases that show sentiments, such as "excellent," "satisfied," and "impressive," while the negative lexicon includes terms like "poor," "disappointed," and "frustrating,".

2.1.4 VADER Classifier

The last lexicon-based approach that we tested is VADER (Valence Aware Dictionary and Sentiment Reasoner) [5]. It employs a pre-established lexicon containing over 7500 words and phrases. Each paired sentiment intensity score that ranges from -4(highly negative) to +4(highly positive). VADER classifier analyzes text by breaking it into individual words, assigning sentiment scores to each, and then calculating the overall score. Its rule-based features enhance sentiment analysis by effectively handling negation and incorporating grammatical conventions. For instance, it reverses sentiment scores for phrases that are negated, such as "not good," and modifies scores according to intensity boosters, like "very" or "barely." Furthermore, VADER takes punctuation and capitalization into account to enhance the accuracy of sentiment assessments.

2.2 Lexicon-based classification

The classification method we employed for the lexicon-based classifier was simple and straightforward. In this approach, each new bug is classified into the category that has the highest overlap with its associated words from the lexicon. Specifically, we determined this overlap by identifying which category's lexicon contained the most matching words. Importantly, while we calculated ratios to assess word relevance, these ratios were not used as weights in the classification process. Instead, they served solely as a means to set thresholds that limited the number of words considered for classification.

$$\text{Predicted Severity} = \begin{cases} \text{Severe}, & \text{if } |B \cap S| > |B \cap NS| \\ \text{Non-Severe}, & \text{if } |B \cap NS| > |B \cap S| \end{cases} \tag{1}$$

where B stands for the set of words in the new bug report and S stands for the set of words in the Severe lexicon while NS denotes the set of words in the Non-Severe lexicon. In cases where a new bug does not fall under the conditions specified in equation 1, i.e., the intersection of words between the bug and the severe and non-severe lexicons is equal, we considered these bugs to be neutral. To further classify these neutral bugs, we examined the ratios of the severe and non-severe word. If the ratios were exactly equal and are non-zero, we iterated through each word in the lexicon and found its index in the severe lexicon, we calculated the normalized indices of the word by dividing its index in the severe dictionary by the total length of the severe dictionary and multiplying by 100. These normalized indices are stored in a list. We repeat this process for the non-severe dictionary, creating a separate list of normalized indices for the non-severe words.

$$\text{Severe Word normalized indices} = \frac{I_{sl}}{L_{sl}} \times 100 \tag{2}$$

$$\text{Non-Severe Word normalized indices} = \frac{I_{nsl}}{L_{nsl}} \times 100 \tag{3}$$

where I_{sl} denotes the index for severe lexicon and L_{sl} denotes the length of the severe lexicon. I_{nsl} denotes the index for non-severe lexicon and L_{nsl}denotes the length of the non-severe lexicon.

Next, we created two separate data frames: one for the severe normalized indices and another for the non-severe normalized indices. From these, we calculated the minimum value of the normalized indices for each category. This step was crucial in helping us identify the category with the lowest overall normalized indices, which we hypothesized to be the most representative of the bug report's severity level. Following this, we compared the minimum normalized indices for both categories. The category with the lower normalized indices was then identified as the predicted severity level for the bug report.

In cases of neutral bugs, where the intersection of words between the bug report and the severe and non-severe lexicons was equal, and the ratios of severe and non-severe word normalized indices were exactly equal to zero, we employed a random assignment approach to determine the severity level. Specifically, we randomly assigned the bug report to either the severe or non-severe category.

It is important to note that this random assignment approach was utilized as a last resort when all other lexicon-based classification methods had been exhausted. Moreover, this random assignment was only applied to a small subset of neutral bugs. By incorporating this strategy for zero-equal ratio neutral bugs, we aimed to provide a comprehensive and robust lexicon-based approach for bug severity classification. Ultimately, this method ensured that all bug reports were assigned a severity level, even in cases where the textual content alone did not provide a clear indication of the bug's criticality.

3 Empirical Evaluation

3.1 Data

The data used in this study was taken from Bugzilla and Eclipse online platforms, comprising bug reports that include attributes such as Bug ID, Product, Component, Assignee, Status, Resolution, Summary, Changed On, Priority, Severity, and Type. The primary focus was on the target Severity column and the Summary column, which provided unstructured textual descriptions of the bug in the bug report. Severity levels are categorized as blocker, critical, major, normal, trivial, and minor (or S1 to S4), with the Type column indicating whether the issue is a defect, enhancement, or task. After combining data from both software projects and excluding entries tagged as N/A, the dataset consisted of 51,154 items, with 33,924 classified as non-severe and 8,214 as severe. The Summary column was pre-processed with tokenization, stop words removal and lemmatization.

3.2 Experimental Setting

In this section, we discuss the new experiments conducted for static lexicon creation and the metrics selected to evaluate the models. The F1 score serves as a crucial metric for assessing the performance of classification models, particularly in our context of dealing with imbalanced datasets. It combines precision and recall into a single score, providing a balanced view of both metrics. These metrics were chosen for comparison due to their ability to highlight the trade-offs between false positives and false negatives, which is essential given the imbalanced nature of our dataset. A series of experiments were undertaken to generate static lexicons for software bug severity classification.

The first group of experiments (Exp 1a,1b,1c,1d) utilized the threshold-based approach, where a lexicon was created in the first experiment (1a), using a dataset from Firefox for training while testing was performed using a dataset from Eclipse shown in Table 1, and a static lexicon was created named **lexicon_THR_1a**. For second experiment (1b) from same group of experiments, the datasets were switched: using a dataset from Eclipse for training while testing was performed using a dataset from Firefox shown in Table 2. The lexicon created from experiment 1b is named as **lexicon_THR_1b**. These two lexicons were further merged by removal of duplicates and the final lexicon was generated named **lexicon_THR_1ab**, which included 1231 severe words, 6719 non-severe words. In the same group of study, third experiment was conducted with combined dataset of Eclipse and Firefox split into 60|20|20. A new lexicon generated from this experiment was named **lexicon_THR_1c**, which comprised 2130 severe words and 5479 non-severe words. The forth experiment within this group focused on the combined dataset but filtered out bugs having a normal status to ensure equal data distribution. Again, the data was split with a ratio of 60|20|20, resulting in another static lexicon named **lexicon_THR_1d**, which included 45 severe words and 764 non-severe words.

In the second group of experiments (2a,2b), we employed a linear SVM coefficient approach to generate the lexicon. Initially, the lexicon named **lexicon_SVM_2a** was created using the Firefox dataset, and tested using the Eclipse dataset shown in Table 1. Subsequently, another lexicon named **lexicon_SVM_2b** was created from the Eclipse dataset and tested on the Firefox dataset also shown in Table 2. Words with positive coefficients indicated the severe category, while words with negative coefficients indicated the non-severe category. This way, two static lexicons were created, merged, and duplicates were removed based on coefficient values, resulting in a final static lexicon named **lexicon_SVM_2ab**, which consists 3,756 severe words and 7,307 non-severe words.

In the third group of experiments, we utilized publicly available lexicons such as Bing Liu's opinion lexicon (lexicon_BL) to classify bug reports from both Eclipse and Firefox. In the last experiment from this group, we utilized the VADER rule-based classifier to classify bugs from both Eclipse and Firefox bug reports. These experiments were conducted to compare the results of our generated lexicons from two different approaches against existing lexicons.

4 Results

4.1 Performance Assessment

Results of our initial experimental assessment for dynamically created lexicons were already published [1]. According to F1 score for the Severe category as the most relevant performance measure in our setting, the lexicon approach looked promising in comparison to the machine learning approaches. Here we present the results of the new experiments conducted with static lexicons. In this study, two datasets, Firefox and Eclipse, were employed to evaluate various approaches for predicting software bug severity. The methodologies examined included threshold-based lexicon creation, linear SVM coefficient-based lexicon generation, and the utilization of publicly available lexicons (lexicon_BL and VADER). The results of the approaches studied in this work are presented for the Eclipse and Firefox datasets.

Results from the experiments where Eclipse is used as a testing dataset are shown in Table 1. They indicate that all the approaches in this setting performed better than the

Table 1: Results showing averages of 10 iterations of experiments where static lexicon was created with Firefox dataset and tested with Eclipse dataset

	F1	F1mean	TP	FP	TN	FN	tl	tc
lexicon_THR	0.2208	0.4137	1147.5	2625.3	18212.7	5407.5	609.8	20.5
lexicon_SVM	0.3247	0.4670	2609.4	6883.7	13954.3	3945.6	182.4	29.6
lexicon_BL	0.3540	0.4806	4065.8	12348.1	8489.9	2489.2	-	16.4
Vader	0.2831	0.5144	2031	5758	15080	4524	-	130.2
SVM	0.4705	0.5454	321.8	612.1	20225.9	6233.2	501.1	847.5
MNB	0.4782	0.5456	390.3	697	20141	6164.7	1.5	0.7
LG	0.4418	0.5480	201.6	521.3	12564.7	1457.4	5.3	0.6

Table 2: Results showing averages of 10 iterations of experiments where static lexicon was created with Eclipse dataset and tested with Firefox dataset.

	F1	F1mean	TP	FP	TN	FN	tl	tc
lexicon_THR	0.2800	0.4634	688.2	2559.2	10526.8	970.8	825.3	7.3
lexicon_SVM	0.2362	0.4405	640.3	3098	9988	1018.7	6973.8	34.7
lexicon_BL	0.2141	0.4297	1132.5	7785.4	5300.6	526.5	-	8.9
Vader	0.2218	0.5112	667	3688	9398	992	-	70.6
SVM	0.1805	0.5454	262.7	984	12102	1396.3	4120.5	665.3
MNB	0.1727	0.5456	228.8	758.9	12327.1	1430.2	3.55	0.3
LG	0.1691	0.5480	201.6	521.3	12564.7	1457.4	15.9	0.3

classifier using our threshold-based lexicon approach. However, computing time requirements of our threshold-based classifier are lower than for majority of the other approaches.

Comparative analysis of F1 scores across all methods for the Eclipse dataset reveals some insights into performance of various classification techniques. SVM recorded F1 score of 0.47 which is comparable to Multinomial Naive Bayes (MNB) with F1 score 0.48. Logistic Regression (LG) with an F1 score of 0.44, showed reasonable performance. The Linear SVM-based lexicon (lexicon_SVM) recorded a much lower F1 score of 0.32, The opinion lexicon-based (lexicon_BL) approach achieved F1 score of 0.35. The Vader approach achieved F1 score of 0.28 and lexicon_THR recorded the lowest F1 score of 0.22. These results suggest that in this setting the lexicon-based approaches are less effective compared to classic machine learning techniques.

In comparison of computing time requirements of classification, the results indicate that the machine learning models, MNB and LG, demonstrated good efficiency with processing times of 0.7 seconds and 0.6 seconds, respectively. In contrast, the lexicon_BL required 16.4 seconds, while our lexicon-based approach took 20.5 seconds. The maximum computational time was recorded for the SVM at 847.5 seconds, followed by VADER at 130.2 seconds and lexicon_SVM at 29.6 seconds.

Results from the study in which the lexicon was created with Eclipse dataset and tested on the Firefox dataset are shown in table 2 In the Firefox dataset, the evaluation of various classification methods based on their F1 scores indicates all the approaches achieved a very low F1 score; however, lexicon_THR with an F1 score of 0.28, performed better than all the other approaches. The SVM approach achieved F1 score of 0.18, which is relatively low. MNB achieved F1 score of 0.17. LG yields F1 score of 0.17. lexicon_SVM performed better than SVM and MNB with an F1 score of 0.24. lexicon_BL achieved an F1 score of 0.21 and showed moderate performance. The Vader rule-based lexicon approach recorded an F1 score of 0.22.

We compared the computing time of individual classifiers and results show that MNB and LG had the best readings of 0.3 seconds. This was followed by lexicon_THR which spent 7.3 seconds and lexicon_BL that performed well with 8.9 s. lexicon_SVM took 34.7 seconds and Vader took 70.6 seconds. SVM classification took by far the longest with 665.3 s. These results demonstrate that the lexicon-based approaches are efficient, but not as much as some of the simpler classic machine learning methods, while the VADER classifier tends to be less efficient and particularly the SVM-based machine learned classifier is in a class of its own regarding this aspect.

4.2 Qualitative Analysis

The purpose of performing a qualitative assessment in this work is to analyse how the choice of words in the lexicons is related to the severity of software bugs. In few cases, a negative term may be classified in the non-severe category, or conversely, a positive term may be assigned to the severe category as a result of the specific vocabulary employed by the individual bug reporter. For this purpose the top 25 words from the aforementioned approaches are being enlisted below.

4.2.1 Intersection of lexicon_BL with Eclipse dataset

The purpose of the qualitative analysis regarding lexical intersections of lexicon_BL and the summary text column in software bug report for Eclipse is to explore the relationship between the languages used by users in bug reports and the corresponding lexical choices made in the opinion lexicon. The aim is to identify patterns that may help in categorising the bug severity correctly. Interestingly, lexicon_BL performed comparatively better on the Eclipse dataset, which makes it interesting for a qualitative comparison in this case. The methodology for determining intersection is an exact match of words from lexicon_BL and the words from the summary column in the bug report. The most frequently intersecting examples are provided below:

Intersection of lexicon_BL positive words with Eclipse bug report summary: work: 494, support: 329, properly: 221, correctly: 214, available: 133, dynamic: 107, correct: 74, supported: 70, right: 61, improve: 57, respect: 53, better: 52, refresh: 49, clean: 47, like: 42, improvements: 40, proper: 39, consistent: 35, progress: 35, well: 32, leads: 32, top: 29, works: 27, refreshing: 24, contribution: 23

Intersection of lexicon_BL negative words with Eclipse bug report summary error: 1001, fails: 461, errors: 377, incorrect: 274, wrong: 248, problem: 216, invalid: 202, problems: 184, broken: 182, warning: 164, incorrectly: 153, hangs: 108, failure: 105, issues: 102, bug: 93, disabled: 86, fail: 86, failed: 86, object: 86, failing: 83, slow: 83, static: 80, failures: 76, deadlock: 71, breaks: 69

Category-wise list of the intersecting words with the lexicon_BL positive words

Words only in severe category: prefers, rapid, protection, led, intelligent, guidance, success, unmatched, overtaken, reliably, succeed, favorite, responsive, recovery, satisfied, simplified, eagerly, warm, positives

Words only in non-severe category: restructure, positive, readable, helpful, variety, fairly, enhance, faster, pretty, colorful, smooth, important, interesting, precise, trivially, hot, recommended, compact, integrated, enough, refinement, welcome, nice, clearly, simplest

Category-wise list of the intersecting words with lexicon_BL negative words

Words only in severe category: radically, incapable, troubling, miss, messed, unfit, destroy, inadequate, erase, killing, discriminate, difficulty, flawed, unsafe, refuse, cripple, dubious,

repetitive, die, inappropriately, stealing, erratically, pan, painful, meaningless, zombie

Words only in non-severe category: obscures, clash, irreversible, omit, confusion, insensitively, severity, ugly, harmful, harsh, leakage, idle, limited, redundancy, limits, flicker, inaccurate, annoying, oddities, bleeding, trivial, lag, bleed, mangled, distortion, trapped

The intersecting words from lexicon_BL positive and negative word lists and the bug report fall in both severe and non-severe categories. In the severe category, words like "incapable," "troubling," "flawed", "destroy", "kill", "painful", and "difficulty" express strong negative sentiments, indicating issues that could impact user experience and are correctly classified. However, some words, like "zombie" and "messed," , succeed", and "favorite" may lack clarity and seem to be misclassified.

In the non severe category words like "helpful" and "readable", "confusion," "annoying," "ugly," trivial," "idle," and "limited" indicate minor issues and accurately suggest enhancements and are correctly classified. Interestingly, words like "harmful" and "irreversible," while serious in some contexts, seem to be misclassified as non-severe.

4.2.2 Intersection of threshold-based custom lexicon with Eclipse dataset

Intersecting Severe words from threshold approach with Eclipse bug report:

not: 4470, file: 1178, error: 1001, server: 870, wizard: 761, new: 581, web: 564, work: 494, missing: 477, fails: 461, build: 426, version: 410, update: 361, service: 344, support: 329, create: 276, message: 269, add: 268, extension: 264, model: 260, plugin: 257, open: 249, creating: 248, change: 245

Intersecting Non-severe words from threshold approach with Eclipse bug report: not: 4470, project: 1627, file: 1178, editor: 795, validation: 770, page: 725, new: 581, view: 572, web: 564, work: 494, missing: 477, name: 456, type: 454, source: 438, content: 424, version: 410, class: 388, default: 387, attribute: 385, update: 361, runtime: 355, service: 344, exception: 338, code: 337

Category-wise Intersecting Severe words from threshold-based approach

Words only in severe category: ver, succeed, rate, c, erase, refuse, kernel, shift, acknowledge, messed, creat, accessed, reliably, zh, eventually, spanish, weekly, destroy, f, miss, killing, eventual, architecture, cmdline, firing, disconnected, internally, however, mail, cs, ascii, p, garbled

Words only in non-severe category: exiting, username, unstable, indefinite, dropped, uncheck, whitespace, invisible, deactivated, bringing, participant, future, live, excessive, interval, printing, feed, talk, wrapper, attempted, dialogue, together, installable, infinity, depending

Category-wise Intersecting Non-Severe words from threshold-based approach

Words only in severe category: government, reappearing, toolkit, sharing, spike, unintended, inconsistancies, occuring, responsive, accessed, animation, remembers, safety, co, duplication, octal, physical, recorded, reflecting, pickup, oriented, protection, reached, scale, decorated

Words only in non-severe category: numeric, integrate, quoted, visually, zoom, uncheck, significantly, day, cloned, difference, stateless, participant, partly, future, extends, boundary, area, leftover, uniform, buggy, accidentally, licence, unhelpful, stretched, feel, val, pasted

The threshold-based approach classifies words into severe and non-severe bug categories. In the severe category, words like "destroy," "killing," "messed," "refuse," "disconnected," and "inconsistencies" indicate major issues and are correctly classified. However, some words, like "ver," "c," and "zh," lack context and clarity, and seem to be misclassified for the severe category. Conversely, the non-severe category includes words like "excessive", "invisible", "dropped", "attempted" and "together", which indicate minor issues and are correctly classified. Interestingly, words like "buggy" "accidental", "unstable" are classified as non-severe, but imply critical problems, indicating misclassification.

Frequent appearance of "not" suggests the need to consider negating words in the future.

4.2.3 Intersection of Linear SVM based custom lexicon with Eclipse dataset

Intersecting Severe words from Linear SVM approach with Eclipse bug report: not: 4470, error: 1001, server: 870, validation: 770, wizard: 761, new: 581, work: 494, missing: 477, fails: 461, type: 454, build: 426, attribute: 385, runtime: 355,exception: 338, code: 337, module: 329, support: 329, create: 276, incorrect: 274, message: 269, add: 268, extension: 264, model: 260, plugin: 257

Intersecting Non-severe words from Linear SVM approach with Eclipse bug report: project: 1627, file: 1178, editor: 795, page: 725, view: 572, web: 564, name: 456, source: 438, content: 424, version: 410, class: 388, default: 387, update: 361, service: 344, tag: 294, need: 287, dialog: 279, test: 274, schema: 259, set: 251, creating: 248, wrong: 248, change: 245, folder: 244, element: 226

Category-wise Intersecting Severe words from Linear SVM Approach

Words only in severe category: succeed, toolkit, internally, accessed, ascii, architecture, forth, eventually, firing, blocker, unsafe, messed, layer, zh, mail, unpacked, spanish, erase, extreme, killing, ver, garbled, reliably, installer

Words only in non-severe category: partition, pause, attemping, warn, massive, terminates, pretty, nearly, together, whitespace, follows, anywhere, prompting, deactivated, middle, concurrent, designed, draw, unstable, sign, upgraded, phantom, username, ssl, mac, minimize

Category-wise Intersecting Non-Severe words from Linear SVM Approach

Words only in severe category: popups, erratically, encrypted, reached, neither, sharing, crashing, favorite, safety, operational, compose, deactivating, filling, refuse, storing, vulnerability, organized, spike, exporter, decorated, scale, disconnected, responsive, co, reboot

Words only in non-severe category: management, existant, person, significantly, bookmark, accepted, configure, assert, equivalent, ad, supposed, originating, functioning, slowly, dictionary, hiding, fishy, selector, colour, filesystem, frame, busy, automatic, year

In the severe category, words such as "blocker," "unsafe," and "vulnerability", "crashing," "messed," and "killing" show critical issues that could lead to system failure and user frustration. This indicates that these words are correctly classified by the Linear SVM coefficient approach. However, some words, such as "zh," "ver," and "forth," lack clear context and clarity and seem to be misclassified for severe category. On the other hand, the non-severe category includes words like "unstable," "massive," "pretty," "pause," "together," and "whitespace," which indicate minor inconveniences. This shows that these words are correctly classified as non-severe bugs. Interestingly, words like "warn" and "deactivated," seem to be misclassified by the Linear SVM coefficient approach.

4.3 Discussion

The results indicate that the choice of dataset significantly influences the performance of the approaches under study. The threshold-based lexicon and Linear SVM coefficient-based lexicon offer a promising baseline and allow for some transparency as it was shown with the qualitative analysis. The frequency of negation-related words shown by the qualitative analysis indicates the potential importance of handling negation, which we will pursue before the otherwise planned enrichment of the lexicons via vector embeddings. Despite the expectation that the lexicon-based approaches will take the least computation time, the results show that the computation times of lexicon-based classifiers are

in-between the slower ML approaches (Vader and SVM) and the faster ones (MNB and LG), but closer to the latter.

5 Conclusion

Following the initial studies on the potential of a lexicon for classifying bug severity, static lexicons were created using two approaches that combined datasets from two sources. Results indicate that the performance of our threshold-based approach and the publicly available lexicons is dependent on the dataset used. The lexicon-based approach would be more effective if datasets from different domains and well-written bug reports are included in the lexicon creation process. Furthermore, sentiment lexicons proved surprisingly useful, despite their lack of software-specific vocabulary. The analysis also suggests that addressing negation in the lexicon-based approach could enhance its effectiveness.

6 Acknowledgments

This work was partially financed by financial support from the Slovenian Research Agency for research core funding for the programme Knowledge Technologies (No. P2-0103).

References

[1] Fatima Aziz and Martin Žnidaršič. “Potential of lexicon-based bug severity classification”. In: *14th International Conference on Information Technologies and Information Society (ITIS 2023)*. Ljubljana, 2023, pp. 21–28.

[2] Aladdin Baarah et al. “Sentiment-based machine learning and lexicon-based approaches for predicting the severity of bug reports”. In: *Journal of Theoretical and Applied Information Technology* 99.6 (2021).

[3] Anh-Hien Dao and Cheng-Zen Yang. “Severity prediction for bug reports using multi-aspect features: a deep learning approach”. In: *Mathematics* 9.14 (2021), p. 1644.

[4] Minqing Hu and Bing Liu. “Mining and summarizing customer reviews”. In: *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. 2004, pp. 168–177.

[5] Clayton Hutto and Eric Gilbert. “Vader: A parsimonious rule-based model for sentiment analysis of social media text”. In: *Proceedings of the international AAAI conference on web and social media*. Vol. 8. 1. 2014, pp. 216–225.

[6] SM Mazharul Islam, Xin Luna Dong, and Gerard De Melo. “Domain-specific sentiment lexicons induced from labeled documents”. In: *Proceedings of the 28th International Conference on Computational Linguistics*. 2020, pp. 6576–6587.

[7] Tao Zhang et al. “Towards more accurate severity prediction and fixer recommendation of software bugs”. In: *Journal of Systems and Software* 117 (2016), pp. 166–184.

Attributing authorship to short texts on social networks

Ana Vladić, **Juraj Petrović**
University of Zagreb Faculty of Electrical Engineering and Computing
Unska 3, 10000 Zagreb, Croatia
{ana.vladic, juraj.petrovic}@fer.hr

Abstract: *Authorship attribution is a classification task with the goal of attributing authorship of one author from a closed set of authors to a given anonymous text. In context of short texts that are typically linked with social media platforms, attributing authorship can be an important step in forensic analyses and tracking the spread of information online. In this paper we explore the authorship attribution potential of several machine learning models trained and evaluated on two datasets comprised of Twitter posts, as well as how the number of authors, the number of texts, data preprocessing, and data augmentation affect the classification accuracy. While the obtained results suggest that the tested models successfully extracted some knowledge that can be used to correctly assign authorship to short texts, the results, all leave much to be desired due to challenges related especially to the relatively small number of sample texts per author and their length.*

Key Words*: authorship attribution, short texts, tweets, machine learning, social networks*

1 Introduction

Determining the authorship of a text is a task within the field of natural language processing. It involves identifying the author of a text based solely on the text itself, without considering any other potential sources of information about the author. This task, also known as *authorship attribution* or *authorship identification,* can take different forms. In its most common form, it is a classification task with a given set of authors and their known texts based on which one or several anonymous texts should be attributed to one or none of the authors. Other similar tasks include *authorship verification* or *authorship discrimination*, pertaining to the problem of verifying whether two texts were written by the same author, *author profiling*, and *concealment of authorship*. Today, text authorship attribution is important for ensuring that text writers are recognized, protected, and held accountable for their work. In forensic investigations, for example, a single or several short text messages could provide a trace to an author's identity or a link between multiple social media accounts run by the same person or even bots.

Authorship attribution can be attempted on texts of any length, yet the approach to the authorship attribution problem will differ significantly for texts of different lengths, languages, or genres. Longer texts have a number of statistical features that can be used to try to match them with some of the known text samples. Classification of short texts, typically available on social networks, is an especially challenging task due to the large number of users (authors), a limit on post (text) length, and the disbalance in number of training samples per author. Such large-scale authorship attribution is equated in the

literature with looking for a needle in a haystack [1] and cannot rely on common methods of authorship attribution used with long texts.

In the past, data sets of movie reviews IMDb62 [2], posts from blogger.com Blog Authorship Corpus [3], and e-mails of Enron company employees [4] have been frequently used for authorship attribution in short texts. More recently, significant work has been done on models for determining the authorship of comments on the Reddit platform [5] and posts collected from radical Islamist forums on the dark web [6]. Still, among the most commonly used datasets for learning and testing models today are tweets, more specifically tweet datasets by Schwartz et al. [7] and by Rocha et al. [8]. Compared to other types of short text sources on the Internet, Twitter (currently known as X) posts have a more standardized length distribution – previously up to 140 characters, and today up to 280 characters for regular and up to 25,000 characters for premium users. In this paper, we approach the problem of short text authorship attribution focusing on tweets and exploring several machine learning models for their classification.

The remainder of this paper is organized as follows. In section 2, we describe the tweet datasets, data preparation, and properties of machine learning models that we used for the problem of short text authorship attribution problem. Classification results are presented in section 3 and discussed in section 4. Finally, we present our work's conclusions in section 5.

2 Methods

2.1 Dataset

The dataset used in this research [8] contains a total of 10,577,946 tweets from 6,293 different user profiles. The dataset is unbalanced. Along with each publication text, the set also contains grammatical part of speech (*POS*) tags. The maximum text length is 140 characters. The shortest texts in the set contain only one character (typically an emoticon). This dataset, provided by the author, was pre-processed. As indicated by [10], the classification results after masking tags (words with the prefix "#") and references to other users (usernames with the prefix "@") tend be significantly worse, as references to other users indicate a network of users with whom an individual author regularly communicates. The absence of masking of references makes the methods unreliable because during training the model will learn to identify the reference with the author who uses it and later all publications containing the same reference will be incorrectly classified [8]. For this reason, dataset preprocessing includes replacing numbers, references to other users (username with the prefix "@"), tags (including "#" prefix), timestamps, dates, shares (re-tweets), and links within the text with special tags for each of the listed categories (NUM, REF, TAG, TIM, DAT, RT REF, and URL respectively). Numbers, dates, times and links are normalized because they are rare features and it is unlikely that the user will use them in identical form more than once, but the way they are used indicates a particular author [8]. When it comes to determining authorship, there is generally no need for typical text preprocessing steps such as converting uppercase letters to lowercase or stemming words, because differences in the use of elements such as uppercase and lowercase letters, prefixes and suffixes make up the author's style.

In order to remove authors who act like bots and make the classification task easier, the number of duplicates among the posts of each author was analyzed. Authors with a

share of duplicates greater than 10% were removed, resulting in 8,239,445 texts from 4,932 authors left in the dataset.

Shared Twitter posts in this dataset contain the tag "RT REF", followed by the text of the post being forwarded. There are 131,687 such posts in the dataset. While such texts are in the literature typically removed from the subsequent processing [8], [9], we decided to include them, but to remove the text following the "RT REF" tag. The motivation for this decision is the idea that the frequency with which someone else's posts are forwarded can potentially indicate the author.

The dataset, furthermore, contains texts written in different languages. To enhance the results of language detection tools on the texts from the dataset and, finally, improve classification outcomes, the following steps were taken:

- Removed "REF", "RT REF", "TAG", and "URL" masking tags, and replaced each "NUM" masking tag with the number "1"
- Texts that do not contain any letters of any world alphabet are marked as English
- For texts whose length after the previous steps is two characters, it is checked whether they are included in the set of the most common two-letter English words
- Texts containing only one token were checked for variants of frequently used English Internet slang

In order to remove authors with non-English texts from the set, three different Python language detection libraries were used: py3langid, Lingua, and fastText. For each text, the three most likely languages were obtained by each of the three tools. Based on the sum of votes and probabilities, the language was determined as the one with the highest number of votes, unless there is another language with a higher sum of probabilities and more than one vote. Using this process, 85% of the texts in the dataset were classified as texts in English. Due to the possibility of incorrect detection, authors with a share of English texts below 90% were removed from the set.

In the final step, authors with less than 50 texts were removed from the dataset. The result is a set with 6,443,043 texts from 3,677 authors.

In addition to the original texts, using the Python emoji library, a version of the texts was created in which the emoticons were replaced by the "SYM" label in order to assess their influence on the classification.

Aside from the already described dataset preparation needs, there are several more remarks that need to be taken into account regarding the short text datasets or tweets in specific. In order for authorship to be established, the author's style must be unchanging to a sufficient extent or the changes in style must be predictable. Still, this doesn’t have to be the case in practice. For example, an analysis of Twitter posts of 133 authors during a period of 4 years, based on n-gram frequencies, indicated that the vocabulary used by authors changes over time and differently for different authors [11]. Additionally, some authors suggest that almost 30% of tweet authors behave like automated bots and publish texts with an easily recognizable pattern, for example a title and a link to an ad or article [12]. If the problem we are dealing with is to create a model that recognizes and attributes human authorship, the presence of non-human authors must be taken into account because it makes the task seem easier than it actually is in real life and is detrimental to building an efficient classification model. Finally, today, when large language models are used to generate texts that are practically indistinguishable from texts written by humans, the task of successful authorship attribution for short texts might seem impossible or at least limited to cases of well-established short text databases and closed author sets.

2.2 Data encoding for machine learning models

Two different ways of text representation were used to represent short texts in a format suitable for the machine learning models used in this paper:

- **Character n-grams** whose vector representations are learned as part of model learning. An n-gram is a sequence of *n* characters or words from a given text. The advantage of character n-grams is that they can be easily and universally extracted from text in any language. Character n-grams generally capture stylistic patterns in the text that indicate the author, and lexical n-grams indicate the topic of the text [13]. Methods that look at the words of the text give good results in domains where the information about the topic is distinct, such as e-mail, movie reviews and blogs [5]. In this paper, character n-grams are represented according to the example of the paper [9], in which character 4-grams were used for text representation. For n=4, from the set of all 4-grams present in the training set, those that appear in only one text and those that appear in every author are thrown out, thereby reducing the noise and reducing the size of the vocabulary. The length of the n-gram *n* and the minimum number of texts that must contain the n-gram are parameters that can be specified when defining the model. The remaining n-grams are sorted lexicographically and marked with indexes and serve as a vocabulary. The zero index is used to represent n-grams outside the vocabulary. Each text is represented by a sequence of indexes that serve as input to the model.
- **Vector representations of tokens** obtained using a language model learned on a set of English Twitter posts. BERTweet [14] is a language model learned on 850 million English Twitter posts. It is based on the language model BERT [15], yet gives better results than BERT on tasks in the field of natural language processing (grammatical marking, recognition of named entities, text classification) when it comes to posts from Twitter. BERT (*Bidirectional Encoder Representations from Transformers*) is a language model with a transformer architecture trained on Wikipedia articles and literary works. BERTweet model expects an input in form of a text normalized as follows: i) emoticons translated into text using the Python emoji library, ii) references to users masked tagged "@USER", iii) links masked marked "HTTPURL". To perform text normalization, the code available at https://github.com/VinAIResearch/BERTweet was modified to replace tags from the pre-processed dataset [8] with tags required by the BERTweet model. The appropriate tokenizer converts the text into a list of token identifiers. The embedding layer of the model converts each input token into a vector representation that represents that token as an element of the model's vocabulary and its position in the text. The meaning of that token in the context of the whole sentence—which can be different in different sentences—is encoded by the encoder. The attention mechanism in the encoder determines, for each token in a sentence, the extent to which every other token contributes to the understanding of its meaning in the context of that very sentence. The output of the model is a 768-dimensional vector representing a sentence. In this work, the output of the penultimate layer of the model is used, i.e. 768-dimensional vectors representing tokens of the input sentence.

2.3 Machine learning models

All machine learning models evaluated in this research were based on convolutional neural networks, since they make excellent use of signals at the character level (which define an authors' style), can successfully distinguish between a large number of classes, reach decision fast, and perform well on unbalanced datasets [5].

Convolutional neural networks consist of convolutional layers, compression layers, and fully connected layers with a softmax classification output. The output of the convolutional layer is a map of features calculated based on each element and its neighborhood in the previous layer. With compression layers, the network selects the highest value features in each map, thereby selecting the most useful features. If the input to the model is a sequence of character n-grams represented by numerical indices, the models start with a representation module that learns dense d-dimensional vector representations. If the input text tokens are represented by BERTweet vector representations, the representation module is not required. The following three convolutional neural network models were explored in the paper:

- **Parallel convolutional model**, as proposed by [12]. The input to the model is a sequence of character n-grams of length l. The model starts with a representation module that learns dense d-dimensional vector representations (d=300). The result is a $d \times l$ matrix. Feature maps are obtained by one-dimensional convolution using convolutional filters of width 3, 4, and 5. Different widths cover features at the morpheme to word levels. Collapse for each feature map selects the maximum value it contains. This implemented model was implemented for character unigrams, l=140, d=300, and 500 filters.
- **Sequential convolutional model**, as proposed by [9]. This model contains folding layers and k-max compression layers. The folding layer creates dependencies between convolutional filters by adding two consecutive output rows. In the next convolutional layer, the feature in the resulting feature map now depends on the two filters of the lower-level convolutional layer. K-max compression selects the k largest values per dimension.
- **Convolutional model with capsule layer**, as proposed by [13] where a convolutional model with a capsule layer was applied to the text in the form of a series of character unigrams as input. In convolutional networks, in order to achieve shift-invariance, compression layers between two layers convey information about the presence of a feature, but information about spatial relationships between features is lost. Classification is based on the presence of learned features, without considering the relationships between them. In capsule networks, groups of neurons called capsules take spatial relationships into account and learn them using dynamic routing by which capsules in a lower-level layer receive feedback from capsules in a higher-level layer [16]. Each capsule determines the presence of a lower-level feature and its activation parameters, and the output of the capsule is fed to the input of the higher-layer capsule whose feature has the greatest chance of being predicted, and in this way the lower-level capsules learn what to pay attention to, that is, which feature to specialize in [16]. The implemented model has two convolutional layers and one layer of 72 capsules. Each capsule in the capsule layer determines the presence of a feature based on the output of the previous layer. The model with one layer of 72 capsules proved to be the best architecture explored in this paper [13].

3 Results

In order to train, test, and evaluate the three models, each of the dataset used was divided into a subset for learning, validation, and testing in the ratio 0.7:0.2:0.1. Model hyperparameters' values were determined using sparse search over the parameter space. Learning was performed for a maximum of 100 epochs with early stopping based on the loss on the validation set.

3.1 Using n-gram representations

The results of the experiments for 50 or 100 authors with 50 or 100, 1000, and 2000 texts each for the best models are shown in Table 1.

Table 1. Classification accuracy of best models for n-gram representations

	50 authors			100 authors		
Model	50 texts each	1000 texts each	2000 texts each	100 texts each	1000 texts each	2000 texts each
Parallel n=1, d=70, lr=1e-3	0.28	**0.49**	**0.53**	**0.26**	**0.43**	**0.45**
Sequential n=1, d=48, lr=1e-3	0.20	0.41	0.45	0.20	0.34	0.36
Sequential n=4, min_texts=2, d=300, lr=1e-4	0.18	0.41	0.48	0.18	0.36	0.41
Capsules n=1, d=70, lr=1e-4	**0.29**	**0.49**	0.48	0.25	0.41	**0.45**
Capsules n=4, min_texts=2, d=70, lr=1e-4	0.17	0.39	0.46	0.18	0.34	0.39

As expected, all models achieve better results when increasing the number of available training texts. By increasing the number of available texts from 50 or 100 to 1000, the classification accuracy increases approximately twofold and reaches the value of approximately 50%. When fewer texts per author are available for training, the sequential model trained on character unigrams performs better than the same model trained on character 4-grams. By increasing the number of texts, the sequential model trained character 4-grams achieves better accuracy than the same model over character unigrams. These results show, as it was expected, that by learning on a larger number of texts, the model can better detect patterns when it comes to longer strings of characters.

The parallel convolutional model consistently achieves the best results, and the sequential one the worst results. Based on the obtained results, the parallel convolutional model most successfully attributes authorship to Twitter posts from the dataset [8].

3.2 Using BERTweet representations

In order to analyze the impact of using BERTweet representations on classification accuracy, experiments were conducted with the same context as with n-grams and the results for the best obtained models are shown in the Table 2.

Table 2. Classification accuracy of best models for BERTweet representations

	50 authors			100 authors		
Model	50 texts each	1000 texts each	2000 texts each	100 texts each	1000 texts each	2000 texts each
BERT_parallel, lr = 1e-4	**0.44**	**0.57**	**0.60**	**0.38**	**0.50**	**0.51**
BERT_sequential, lr = 1e-4	0.36	0.48	0.52	0.29	0.41	0.44
BERT_capsules, lr = 1e-4	0.35	0.54	0.58	0.32	0.46	0.48

Models trained on BERTweet representations perform better than those trained on n-gram representations (Table 1). Although the BERTweet language model was trained for different tasks in the field of natural language processing, it was trained on a huge dataset. Vector representations of character n-grams are learned for this specific problem, but on several orders of magnitude smaller data sets and this is a possible explanation for why they give worse results.

Some additional insights into classification results confirmed initial assumptions about the effects of lower text length on classification outcomes. For example, classification outcomes for authors who write shorter texts are significantly worse compared to classification outcomes of authors of longer texts. *BERT_parallel* model trained on 50 authors with 1000 texts each classified correctly 100% of test examples for the author 4557, whose short texts include the following examples:

- *One third of all recorded burglaries are against businesses: Did you know that one third of all recorded burgl... URL*
- *We're looking for: homes with a stylish family kitchen: Three beautiful British homes for sale with kitchens y... URL*
- *Easter NUM: three family homes with chicken coops: Three gorgeous family homes with chicken coops. URL*

yet the same model classified correctly only 38% of test examples for the author 4557, whose short texts with inconsistent capitalization and informal language include the following examples:

- *REF u ok*
- *REF u think THIS is A JOKE*
- *Im bored so there for i will sleep. TAG*
- *IM SO DUMB*

3.3 Data augmentation

With data augmentation, the dataset size is artificially increased without collecting new data samples. If the task at hand is to classify more than one unknown text, but all belonging to the same author, by combining several texts into one longer text the model has more information available during learning and prediction on the basis of which it can make decisions. By concatenating any two texts from a single author, the number of texts increases, for example, from 50 to 50×49=2450. The experiments were conducted

on data from the set [8] for 50 authors with 50 texts and 50 authors with 1000 texts. The results are presented in Table 3 and Table 4. Each tested model achieves better classification accuracy on augmented texts. This strategy reduces the impact of noise and thereby achieves regularization of the model [9]. The model in each epoch sees each data in several different versions and has the possibility to correct the mistakes made earlier.

Table 3. Classification accuracy of best models for n-gram representations with and without data augmentation for 50 authors with 50 texts

Model	No augmentation	Augmentation
Parallel n=1, d=70, lr=1e-3	0.28	**0.36**
Sequential n=1, d=48, lr=1e-3	0.20	**0.29**
Sequential n=4, min_texts=2, d=300, lr=1e-4	0.18	**0.27**
Capsules n=1, d=70, lr=1e-4	0.29	**0.36**
Capsules n=4, min_texts=2, d=70, lr=1e-4	0.17	**0.31**

Table 4. Classification accuracy of best models for BERTweet representations with and without data augmentation for 50 authors with 50 texts

Model	No augmentation	Augmentation
BERT_parallel, lr = 1e-4	0.44	**0.61**
BERT_ sequential, lr = 1e-4	0.36	**0.50**
BERT_capsules, lr = 1e-4	0.35	**0.58**

4 Discussion

The results obtained through experiments and described in the previous paper section are in line with our expectations. Classification results obtained on a sample of 50 authors with 50 texts per author for three different convolutional neural network architectures have accuracy between 17% and 29% (for n-gram representations) and 35% and 44% (for BERTweet representations). Increasing the number of texts per author to 2000 improved those results to 45% to 53% and 52% to 60% respectively. It seems, based on this, that BERTwitter representations are a more efficient way of encoding text for the authorship attribution of short texts. Among the tested models, parallel convolutional model proved to be the most successful across nearly all experimental settings and the explored data augmentation technique of text concatenation provided an additional increase in classification accuracy across all settings.

Overall, however, no model in no settings achieved results better than 61% in terms of classification accuracy suggesting the task is still challenging due to its constrains – a high potential number of authors and low amount of information available per short text. Due to the same constraints, it cannot be claimed that the results would generalize well on short texts obtained through similar yet different social platforms like Facebook or Instagram. Different platforms have different post length limits, as well as specific limits on post or comment preview length, which authors take into consideration when posting. We believe that the solution to this problem is more likely to be found by training models specifically for a given platform and taking into account its limitations then by training general classification models with input text samples of diverse limitations and features.

5 Conclusion

The work described in this paper describes the challenges and the experimental results related to short text authorship attribution using convolutional neural network classifiers. Several experiments conducted on a preprocessed tweets dataset [8] indicated the complexity of preprocessing required to obtain a dataset with as little noise as possible as well as how challenging it is to obtain high accuracy results on such a dataset. The conducted experiments indicated that BERTweet representations are a better choice than n-grams, and that a parallel convolutional model is a better choice than sequential model or using capsules. Still even with those choices, the task remains challenging and the highest obtained classification results do not exceed 61% accuracy.

6 References

[1] M. Koppel, J. Schler, and S. Argamon, “Computational methods in authorship attribution,” *Journal of the American Society for Information Science and Technology*, vol. 60, no. 1, pp. 9–26, 2009.

[2] Y. Seroussi, I. Zukerman, and F. Bohnert, “Collaborative Inference of Sentiments from Texts,” in *User Modeling, Adaptation, and Personalization*, P. De Bra, A. Kobsa, and D. Chin, Eds., Berlin, Heidelberg: Springer, 2010, pp. 195–206. doi: 10.1007/978-3-642-13470-8_19.

[3] J. Schler, M. Koppel, S. Argamon, and J. Pennebaker, “Effects of Age and Gender on Blogging.,” Jan. 2006, pp. 199–205.

[4] B. Klimt and Y. Yang, “The enron corpus: a new dataset for email classification research,” in *Proceedings of the 15th European Conference on Machine Learning*, in ECML’04. Berlin, Heidelberg: Springer-Verlag, Sep. 2004, pp. 217–226. doi: 10.1007/978-3-540-30115-8_22.

[5] S. Ruder, P. Ghaffari, and J. Breslin, “Character-level and Multi-channel Convolutional Neural Networks for Large-scale Authorship Attribution,” *ArXiv*, Sep. 2016, Accessed: Sep. 30, 2024. [Online]. Available: https://www.semanticscholar.org/paper/Character-level-and-Multi-channel-Convolutional-for-Ruder-Ghaffari/3aea08d079419ae34149ba344df4c092819223b5

[6] L. Ranaldi, F. Ranaldi, F. Fallucchi, and F. M. Zanzotto, “Shedding Light on the Dark Web: Authorship Attribution in Radical Forums,” *Information*, vol. 13, no.

9, Art. no. 9, Sep. 2022, doi: 10.3390/info13090435.
[7] R. Schwartz, O. Tsur, A. Rappoport, and M. Koppel, “Authorship Attribution of Micro-Messages,” in *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, D. Yarowsky, T. Baldwin, A. Korhonen, K. Livescu, and S. Bethard, Eds., Seattle, Washington, USA: Association for Computational Linguistics, Oct. 2013, pp. 1880–1891. Accessed: Sep. 30, 2024. [Online]. Available: https://aclanthology.org/D13-1193
[8] A. Rocha *et al.*, “Authorship Attribution for Social Media Forensics,” *IEEE Trans.Inform.Forensic Secur.*, vol. 12, no. 1, pp. 5–33, Jan. 2017, doi: 10.1109/TIFS.2016.2603960.
[9] A. Theophilo, R. Giot, and A. Rocha, “Authorship Attribution of Social Media Messages,” *IEEE Transactions on Computational Social Systems*, vol. PP, pp. 1–14, Nov. 2021, doi: 10.1109/TCSS.2021.3123895.
[10] R. Layton, P. Watters, and R. Dazeley, “Authorship Attribution for Twitter in 140 Characters or Less,” in *Proceedings of the 2010 Second Cybercrime and Trustworthy Computing Workshop*, in CTC ’10. USA: IEEE Computer Society, Jul. 2010, pp. 1–8. doi: 10.1109/CTC.2010.17.
[11] H. Azarbonyad, M. Dehghani, M. Marx, and J. Kamps, “Time-Aware Authorship Attribution for Short Text Streams,” in *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, in SIGIR ’15. New York, NY, USA: Association for Computing Machinery, Aug. 2015, pp. 727–730. doi: 10.1145/2766462.2767799.
[12] P. Shrestha, S. Sierra, F. González, M. Montes, P. Rosso, and T. Solorio, “Convolutional Neural Networks for Authorship Attribution of Short Texts,” in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, M. Lapata, P. Blunsom, and A. Koller, Eds., Valencia, Spain: Association for Computational Linguistics, Apr. 2017, pp. 669–674. Accessed: Sep. 30, 2024. [Online]. Available: https://aclanthology.org/E17-2106
[13] C. Suman, A. Raj, S. Saha, and P. Bhattacharyya, “Authorship Attribution of Microtext Using Capsule Networks,” *IEEE Transactions on Computational Social Systems*, vol. PP, pp. 1–10, Apr. 2021, doi: 10.1109/TCSS.2021.3067736.
[14] D. Q. Nguyen, T. Vu, and A. Tuan Nguyen, “BERTweet: A pre-trained language model for English Tweets,” in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, Q. Liu and D. Schlangen, Eds., Online: Association for Computational Linguistics, Oct. 2020, pp. 9–14. doi: 10.18653/v1/2020.emnlp-demos.2.
[15] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, J. Burstein, C. Doran, and T. Solorio, Eds., Minneapolis, Minnesota: Association for Computational Linguistics, Jun. 2019, pp. 4171–4186. doi: 10.18653/v1/N19-1423.
[16] A. Kumar, “Why Do Capsule Networks Work Better Than Convolutional Neural Networks?,” Medium. Accessed: Sep. 30, 2024. [Online]. Available: https://ashukumar27.medium.com/why-do-capsule-networks-work-better-than-convolutional-neural-networks-f4a105a53aff

Anthropology and Algorithms: The Good, the Bad, and Entangled

Ana Bezić
anabezic@gmail.com

Abstracts: *As everyday life becomes increasingly mediated by data and algorithms, anthropology offers a critical lens to examine how algorithms intersect with cultural practices, social justice, and power dynamics. Can algorithmic predictions or our digital practices truly account for the unpredictability and complexity of human experience? What rights do we have concerning data rights in a world governed by these systems? This paper explores the anthropological study of algorithms as and of culture—unstable 'objects' that are enacted through the varied practices of those who engage with them. I will discuss how anthropological methods and concepts can be employed to make sense of the precarious relationship with algorithms, what implications these technologies have on our everyday life, and what does anthropology bring to the table in understanding them. By bringing an anthropological perspective to the study of algorithms, this paper aims to entangle the ways in which they influence and reflect societal norms, providing deeper insights into the human impact of these technologies.*

Key Words: anthropology, algorithms,

1 Introduction

Anthropology is a very wide term that encompasses the study of everything of what makes us human. From our origins and biological features to our everyday practices. To understand this complexity of what we all share as humans and our rich diversity, anthropology draws and builds upon knowledge gained from multiple disciplines from the social and biological sciences as well as the humanities and physical sciences. In the last decade, anthropology has opened itself to include a multitude of other than human beings and have begun asking questions on the diversity of ways these non-human beings feature in human lives (Haraway 2008; Konh 2013; Tsing 2013). Computational algorithms are neither. They are a key concept in digital technologies, briefly described as the instructions or rules written in an appropriate language for a computer to process a task and, subsequently, to perform and execute it. Digital technologies in general and algorithms in particular have conditioned a creation of another kind of being, of being between or beyond human and non-human, coinciding neither, affecting both.

Everyday life became increasingly mediated by data and algorithms. While they take life of their own, they are also taking our lives so to speak. Anthropology offers a critical lens to examine this newly founded relations and how these objects intersect with cultural practices, social justice, and power dynamics. Can algorithmic predictions or our digital practices truly account for the unpredictability and complexity of human experience? What rights do we have concerning data rights in a world governed by these systems? This paper explores the anthropological study of algorithms as unstable 'objects' that are enacted through the varied practices of those who engage with them. Anthropology, as I aim to show, have been involved in studying algorithms in different contexts revealing how algorithms, not only influence and are influenced by larger social, historical and political dynamics but have become so entangled with our lives that we have become data. I will also present new and emerging directions, away from anthropological concerns of algorithmic effect and affect on humans towards exploration of the life of algorithms themselves, that define, refine and produce horizons of being in the contemporary world.

2 Anthropology and algorithms

Anthropology has been exploring the topic of algorithms for some time now. This interest emerged alongside the rise of digital technologies and internet in the 1990s and got amplified with the rise of social media in 2010. However, when social scientists speak about algorithms, they tend to be less concerned with the mechanical term, and more with the ways in which software conditions our very existence (Couldry and Mejias 2019).

Anthropologists studying algorithms work in a variety of field sites to understand how algorithms shape social, cultural, and economic life. Together, these works offer critical insights into how algorithms produce effect and affect. By saying anthropology study algorithms, it is unclear whether they refer to the same things. It is a place for imaginary (Bucher 2017), anxieties (de Vries and Schinkel 2019), as culture and systems (Seaver 2017), ecologies (Schinkel 2023) and assemblages (Lowrie 2018). It is undeniable that algorithms do things to us. They shape experiences without needing direct understanding of their technical details. Heterogeneous entanglements we have with algorithms are mostly characterised by their opacity and elusiveness (Pasquale 2015) and by being black boxed they become complex ethnographic objects of concern.

At the beginning, algorithms were simple. For example in music recommendation system they simply picked one artists and seed to similar artists. Now, there are sub algorithms trying to determine what kind of listener you are and they adjust recommendations accordingly. And the master algorithm that binds all this together is optimised to doing one thing only, to keep the listeners listening. From a phenomenological viewpoint, inspired by Merleau-Ponty (1974), algorithms in most of the anthropological studies, are not directly perceived as their mathematical structures, but rather through the moods, emotions, and sensations they generate. Just as we perceive people through their habits or personalities rather than their physical appearance, algorithms are "felt" through the outcomes they produce, not their internal workings.

Early writings have laid the groundwork for questioning the neutrality and autonomy of computational systems and understanding how these systems interact with social structures and identities (Suchman 2007, Dourish 2004). Suchman in her study (1987) investigated relationship between humans and machines and introduced the notion of interactive artefacts to challenge the idea that these can autonomously interpret and act upon user input. The meeting of the human and the machine, each figured to possess pre-established capacities, is in her writing shifted to the possibility of encounter that Suchman understands as ‘those moments of moving complicity between persons and things’ (245). This observation will become a recurring theme in understanding human/posthuman interaction, the amount of complicity and, when it comes to illusiveness of algorithms, our entrapment.

The early writings have paved the way for how social contexts and human embodiment shape algorithms and computational processes acknowledging their embeddedness in a broader social and physical environment (Dourish 2004). Some authors have very early on been critical of classification systems and standardisation that produce obscurity as to what voices and which concerns have been included and/or excluded (Bowker and Leigh Star 2000). Many continue to be sceptical to the possibility of making all knowledge retrievable and re-debatable and to what purpose (ibid: 30). Digital governance has become a corner stone of anthropological inquiry benefitting from work of Gillespie (2018) on surveillance, who has examined how society’s

organisation is influences by automated systems, and Costello et al. (2023) whose research is particularly concerned on public health and youth well being. While there is a discernible pattern of algorithms as social artefacts influenced by and influencing societal structures the exact way these show varies depending on the context. Austin, for example, was able to show how automated decision making systems can perpetuate existing inequalities and shape individual behaviour and how such systems are in need for policy interventions that address the risk posed by social media algorithms to adolescent mental health (ibid).

More recently, anthropologists stepped into various virtual field sites where they begun to understand the ways in which algorithms shape content visibility and user engagement. While researching Spotify and Netflix, Seaver (2022) questions were directed towards understanding how algorithms interpret and predict tastes. Not only they, the algorithms, recommend music, they also create and reinforce specific music taste. That algorithms influence people to adapt their lives to algorithmic expectations and commodify human behaviour comes no longer as a surprise. At the same time they continue to being seen as responsible to contributing to systemic injustices and harbour biases that disadvantage marginalised groups as algorithmic presumed neutrality continues to produce "algorithmically sustained discrimination" (Benjamin 2019: 143).

3 Algorithms as everyday life

We often think of algorithms as neutral, objective, or purely technical—designed to help solve complex problems efficiently. However, anthropology offers us a different lens. Anthropologists have found that actors enact algorithms differently: engineers for example adjust their code to mediate between the distinctive behaviours of their users while consumers try to work the algorithm as they understood it, to generate more desirable matches and possibilities (Seaver 2022; Devendorf and Goodman 2014).

From the perspective of algorithms as culture, position proposed by Seaver (2022), algorithms are understood not as embedded in but as created by humans and shaped by the social contexts in which their creators operate. However, culture as a term in anthropology have gained scepticism as an explanatory all encompassing concept (Abu-Lughod 1991) because of its implicit holism and homogenising, essentialist tendencies were seemed politically problematic and ill-suited to the conflictual, changing shape of everyday life in anthropological field sites. Like other aspects of culture, algorithms are shaped by practices that merge technical and non-technical aspects, rather than drawing a clear line between the two. In this view, algorithms are not singular technical objects that enter into many different cultural interactions, but are rather unstable objects, enacted by the practices people use to engage with them (Ibid).

When I conducted a search on twitter #algorithm, one user referred to smartphones as "digital pacifiers," noting how algorithms are deliberately crafted to keep users engaged for longer durations. From an anthropological perspective, we can interpret this as a reflection of intentionality embedded in technological artefacts, a concept Alfred Gell explores in *Vogel's Net (1996)*. Gell argues that tools, including algorithms, are not passive objects but are designed with the intent to influence human behaviour, much like a hunter's trap designed to capture prey. The algorithm's purpose here is not just functional but behavioural—it aims to maximise user engagement. Another user raised a nostalgic question: *Can we return to a time of scheduled news, fact-checked newspapers, and in-store shopping?* This speaks to the cultural shift algorithms have brought about,

moving us from passive consumption to algorithmically-driven, personalised interactions. Nick Seaver's (2022) work highlights the intimate and often unnoticed role algorithms have in our digital lives and how they actively shape what we encounter online by prioritising certain content based on what is "engaging" or "viral." This curated experience of media consumption creates a sense of longing for a time when content seemed more neutral and less controlled by algorithmic logic. His work further highlights how users form personal theories about algorithms, of what the algorithm should be, and many express discomfort with the way they are categorised, underscoring the tension between algorithmic categorisation and self-perception despite not fully understanding their technical operations.

The everyday life, where we learn about friendship and love, where we acquire and develop varied competencies, and where potential for our individual and collective agency to transform existing social and political conditions resides, has never been a neutral realm (see Gardiner 2000). Our digital life, now irreversibly entangled with our everyday, and the relationships we are crafting with or despite algorithms become a new spatial and temporal nexus where the tension of subjugation and resistance to institutional strategies to control, shape, and organize are being realised. Anthropologists are on a well known ground and sites to investigate this newly formed precarious existence with algorithms, however complex it may be.

4 Temporality of suspension

As many of us try to navigate the algorithmic perseverance in our daily life, one cannot but feel entrapped by their all pervasiveness. Nick Seaver (2017), borrowing from Alfred Gell's work, further amplifies and shows the ways in which they trap us without us even being aware of it. One can imagine unnecessary cruelty lobsters endure by being cooked alive by being put in a cold water with wine, drunk in a familiar liquid environment not knowing they are being boiled to death. I would like to dwell a bit more at this trap metaphor as it vividly portrays the predicament of this newly created relationship.

Alfred Gell, an anthropologist, compared artefacts to traps in his work *Vogel's Net: Traps as Artworks and Artworks as Traps (1996)*. In the article, Gell critiques conventional definitions of art, proposing that artworks and gallery spaces function as traps. The trap is explored as a particular kind of technology that is not only a model of its creator, but each is also a model of its victim. It is the hunter who knows the victims habitual responses and is able to subvert them. Gell asserts that the tragedy of traps lies in how victims contribute to their own demise through their actions.

Algorithms work like this, as a thought-trap which hold its victim for a time, in suspension. A suspension is a temporary stoppage, not an empty space, rather a space where we, by swiping, liking, and lingering on are seldom aware of the workings of the program that only has one objective, to hold as in suspension for as long as possible. The double bind of power, as Foucault would say (1995), lies in its capacity to both subjugate and empower. As we navigate through social media, suspension becomes a common place to dwell and live, to be subjugated and /or empowered. Can we ever leave it?

This in-between space, could also be conceived as the liminal space, which Turner understood in his ethnography on Ndembu people in Zambia, to be a space where people can experience ambiguity, transformation, and communitas - a sense of unity and equality among participants (1969).

Suspended from social hierarchies, rules, and identities it allows individuals to explore new behaviours, and perspectives. In a Turnerian sense it is a positive and productive space. While it has never been easier to search, find, learn, and communicate, with the help of algorithmic programs we can navigate massive amounts of data, streamline processes, and even make predictive decisions that were once impossible. They can improve healthcare, optimise logistics, and predict future outcomes in ways that save time, resources, and lives. The strength of the algorithm is their efficiency at the speed and at the cost that no conventional human could mach.

Most of the times we do not know all the operations behind the algorithm but what we do know is that while doing these calculations/processes/predictions they often exacerbate social inequalities. Not only that, algorithmic profiling, in everyday life makes people belittle and objectified that is often experienced as a form of violence (Barassi, V. 2023). Many participants in Barassi's study felt bombarded with ads profiled based on their searches - their narratives show the feeling of powerlessness in the construction and representation of their identities. Further complexities emerge when algorithms do not translate across different cultural contexts (Horst and Mohammid 2022). Horst and Mohammid studied Amazon Echo Look, an app that compares outfits and provides fashion recommendations among women in Trinidad - one women complained that the app did not take into account her own preferences, giving recommendation to outfits in which she would appear slimmer.

Another study was conducted among YouTube influencers in London and their perception of algorithms. Glatt (2022) shows how algorithmic invisibility, or work deprioritised on recommendation lists based on algorithmic ranking is a common fear even among more influential YouTubers. Influencers found algorithms to be unpredictable, which underscore the importance of researchers to hold developers accountable for algorithmic processes. Since they rely on existing data to function, any biases in that data can be perpetuated or even magnified by the algorithm. For instance, predictive policing algorithms use historical crime data to forecast where crime might occur. If this data reflects an over-policing of certain neighbourhoods, the algorithm will direct even more resources to those areas, reinforcing cycles of inequality. Jenna Burrell's (2016) work highlights this "opacity" in machine learning—where the decisions and operations of these systems are hidden, making it difficult to understand or challenge their outcomes.

5 The Entanglement, capitalism by algorithmic means

Willem Schinkel, in his work on algorithms, discusses how these systems are not standalone (2023). Rather, they exist within broader social, political, and economic networks—forming what he calls an "algorithmic ecology." This perspective helps us recognise that algorithms are more than technical systems; they are deeply enmeshed in dynamic social processes. For example, social media algorithms do not just reflect users' preferences—they actively shape public discourse. These systems tend to prioritise content that maximises engagement, often promoting the most sensational, emotionally charged material. This can fuel polarisation and even spread misinformation faster than the truth can keep up. Consider recommendation systems like those used by Netflix or Spotify. For example, House of Cards, Netflix tv show, was made based on the searches the users made: based on what it knows about the viewing habits of its 33 million users**.** Every time we engage with these systems, we are engaging with the assumptions embedded in their design (Seaver 2022). We are living in a feedback loop, where algorithms respond to our behaviour,

but at the same time, they influence our actions, perceptions, and social interactions in ways we do not always notice.

If algorithms are shaping our lives, decisions, and access to resources, can we have more control over how our data is used? What rights do we have to know how these systems work, especially when they affect decisions about our future? Many algorithms operate behind a wall of secrecy. While one might presume secrecy to be a simple matter of hiding facts that could be easily revealed, secrecy in practice is not so clear; secrecy is a social process, enacted by both insiders and outsiders (Jones 2014). The recognition of algorithmic secrecy may seams destabilising, and it is difficult to hold an algorithm accountable if we do not know what it is doing. As Daniel Neyland demonstrates through his ethnographic work on an algorithmic accountability, making algorithms accountable often means literally changing them—making them "account-able" (2016). Presumption in most of the anthropological studies is that algorithm are more or less fixed or inscrutable, with essential characteristics. Nyelands follows one particular algorithmic system developed for public transport security and was able to show messy realities of algorithmic implementation, and difficulty in studying it ethnographically.

6 Afterword

What does anthropology bring to the conversation about algorithms? Study of algorithms for anthropologists is a the departure from the idealised image of a fieldworker embedded long-term in a bounded society. Rather than sites, anthropologists conduct ethnographies at an entry points, in coffee places, newspaper articles, and social media. If anthropologists are not stuck with the hidden, what is off limits or intentionally hidden, there is a lot to be done with algorithms. Anthropological interest is not in the specific configuration of a particular algorithm at one moment in time, but in the more persistent cultural worlds algorithms are part of. It helps us see these systems not just as abstract, mathematical models but as social and cultural products. It encourages us to question the decisions embedded within their design and to consider their broader impacts on society.

As algorithms increasingly make decisions that affect everything from the news we consume to the loans we get, we must ask: Are these systems working in our best interest, or are we simply feeding them more data? Are we the data? In T*he Costs of Connection*, Couldry and Mejias (2019) laid ground onto the ways in which this conditioning takes place viewing algorithms as data colonialism, global in ambition they penetrate every aspect of life, both human and non-human. Just as colonial practices once appropriated physical and human resources, modern corporations now extract data form people's everyday activities transforming these into a resource. Daily activities - communicating, shopping, even drinking water - are being captured by data points and turned into profit (2019, xiii). We need to rethink our relationship with algorithms and advocate for systems that respect human dignity and fundamental rights. It is indeed time to "talk to the algorithms" by engaging users as well as the designers, processes, and societal organisations that implement them in daily life.

In this review on anthropology and algorithms (see also Schinkel 2023) it has become clear how intermingled, ubiquitous and entangled have we become in this precarious relationship. Algorithms re-tool and manufacture the very nature of what we term everyday life. As computational culture evolves we can begin to lift underneath or excavate algorithms archaeologically. Meaning, not only

how are they understood and deployed but also how they are being developed and as they are being developed. Also we can begin to research its historicity and the conditions of its possibility, and how have they become mechanisms of control and governance. By treating them archaeologically, we acknowledge that they are product of specific historical, political and institutional framework. Not only understanding what they do, but why, how, and who benefits from them would be a good start. One future direction is looking at the code. It is no longer enough to understand what a code does, but also what it means (Marino 2020). New and emerging critical code studies are paving the way to such enquiries where application of "critical hermeneutics to the interpretation of computer code, program architecture, and documentation within a socio-historical context" can take place (Ibid, 53). Furthermore, by researching socio-technological history of particular codes we can begin to navigate complexities of technological colonialism, imperialism, and misogyny that permeate code and coding culture which are often unacknowledged by its makers. Jenna Ng and David Theo Goldberg developed a Codex, sets of parameters for studying algorithmic determinations and I provide the Codex here in its entirety as a way to start thinking forward towards not a single approach rather as an evolving framework for future research.

CODEX
Data are the currency of our time.
The use and exchange value of data are algorithmically driven.
Data production, collection, distribution, circulation, and ownership operate only as effectively as the algorithms enabling them.
AS is the critical study of the social, political and cultural life of the algorithm and its conditions of change, evolution and possibility. It critically assesses the social order(ing) of life effected by algorithms.
AS is the study of the digitization of data, the data-ification of life, the appification of being, bio-instrumentation and the instrumentalizing and calculability of the everyday.
AS is the study of algorithmic knowledge, its epistemic formations and formulations.
AS is the study of algorithmically produced networks, the machinic production and inter-communication of networked machines. AS is, thus, the study of algorithmically enabled machine life.
AS is the study of the algorithmic subject: the social and political subject produced by and reproducing algorithmic practice, and its modes of being, doing, and becoming.
AS is the study of the algorithmically compelled (pre-)dictable futures, the consuming subjectivities it structures.
AS accordingly is the critical study of the algorithmically modeled restructuring of agency and its loss, autonomy and its strictures and restrictions.
AS, in short, is the critique of algorithmic capitalism, its mode of production, being and subjectivities; of transitory life; of consuming life; of political culture and cultural politics.
AS is the study of the increasing transitivity of data-fied social life, its coding, its codes, and its applications: the algorithmic.

7 References

[1] Abu-Lughod, Lila. "Writing Against Culture." In *Recapturing Anthropology: Working in the Present*, edited by Richard G. Fox, 137–162. Santa Fe, NM: School for Advanced Research, 1991.

[2] N. Costello, R. Sutton, M. Jones, *et al.*, "Algorithms, Addiction, and Adolescent Mental Health: An Interdisciplinary Study to Inform State-Level Policy Action to Protect Youth from the Dangers of Social Media," *American Journal of Law & Medicine*, vol. 49, no. 2-3, pp. 135-172, 2023, doi:10.1017/amj.2023.25

[3] Barassi, Veronica. "Algorithmic Violence in Everyday Life and the Role of Media Anthropology." In *The Routledge Companion to Media Anthropology*, edited by Elisabetta Costa, Patricia G. Lange, Nell Haynes, and Jolynna Sinanan, Routledge, 2023

[4]Benjamin, Ruha. *Captivating Technology: Race, Carceral Technoscience, and Liberatory Imagination in Everyday Life*. Durham, NC: Duke University Press, 2019

[5] Bowker, Geoffrey C., and Susan Leigh Star. *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: MIT Press, 2000

[6] Bucher, Taina. "The Algorithmic Imaginary: Exploring the Ordinary Affects of Facebook Algorithms." *Information, Communication & Society* 20: 130–144, 2017.

[7] Burrell, Jenna. "How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms." *Big Data & Society* 3, no. 1, 2016. https://doi.org/10.1177/2053951715622512.

[8] Couldry, Nick, and Ulises A. Mejias. *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism*. Stanford, CA: Stanford University Press, 2019.

[9] Devendorf, Laura, and Elizabeth Goodman. "The Algorithm Multiple, the Algorithm Material." In
Contours of Algorithmic Life, UC Davis, May 2014.

[10] de Vries Patricia, Schinkel Willem. "Algorithmic anxiety: masks and camouflage in artistic imaginaries of facial recognition algorithms". Big Data Soc 6:1, 2019. https://doi.org/10.1177/2053951719851532

[11] Dourish Paul. Where the Action Is: The Foundations of Embodied Interaction Cambridge, MA: MIT Press, 2004.

[12] Foucault, Michel. *Discipline and Punish: The Birth of the Prison*. Translated by Alan Sheridan. New York: Vintage Books, 1995

[13] Gardiner, Michael. *Critiques of Everyday Life*. London; New York: Routledge, 2000

[14] Gell, Alfred. "Vogel's Net: Traps as Artworks and Artworks as Traps." *Journal of Material Culture*
1, no. 1 (1996): 15–38. https://doi.org/10.1177/135918359600100102.

[15] Gillespie, Tarleton. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. New Haven, CT: Yale University Press, 2018

[16] Glatt, Zane. "Precarity, Discrimination and (In)visibility: An Ethnography of 'the Algorithm' in the
YouTube Influencer Industry." In The Routledge Companion to Media Anthropology, edited by Elisabetta Costa, Patricia G. Lange, Nell Haynes, and Jolynna Sinanan, 544–556. London: Routledge, 2022.

[17] Haraway, Donna J. *When Species Meet*. Minneapolis: University of Minnesota Press, 2008.

[18] Heather A., and Sheba Mohammid. "The Algorithmic Silhouette: New Technologies and the Fashionable Body." In The Routledge Companion to Media Anthropology, edited by Elisabetta Costa, Patricia G. Lange, Nell Haynes, and Jolynna Sinanan, 519–531. London: Routledge, 2022.

[19] Jones, Graham M. "Secrecy." *Annual Review of Anthropology* 43 (2014): 53–69. https://doi.org/
10.1146/annurev-anthro-102313-030143.

[20] Knuth, Donald E. "Seminumerical Algorithms." In *The Art of Computer Programming*, Vol. 2, 3rd
ed., 12–15, 145–146. Boston: Addison-Wesley, 1998.

[21] Kohn, Eduardo. *How Forests Think: Toward an Anthropology Beyond the Human*. Berkeley: University of California Press, 2013.

[22] Lowrie I. "Algorithms and automation: an introduction". Cult. Anthropol. 33:3349–59, 2018.

[23] Marino, Mark. *Critical Code Studies*. New York: The MIT Press, 2020

[24] Merleau-Ponty, Maurice. *Phenomenology of Perception*. Translated by Colin Smith. London: Routledge & Kegan Paul; Humanities Press, 1974. (Originally published in 1945).

[25] Neyland, Daniel. "Bearing Account-Able Witness to the Ethical Algorithmic System." *Science, Technology & Human Values* 41(1): 50–76, 2016. https://doi.org/10.1177/0162243915598056.

[26] Ng, Jenna and David Theo Goldberg "Codex for Algorithmic Studies"
https://algorithmicstudies.uchri.org/codex/ (accessed 3 November 2024)

[27] Pasquale Frank. The Black Box Society. The Secret Algorithms That Control Money and Information. Cambridge, MA: Harvard Univ. Press, 2015.

[28] Schinkel, Willem. "Steps to an Ecology of Algorithms." *Annual Review of Anthropology* 52: 171–186, 2023. https://doi.org/10.1146/annurev-anthro-061421-120931.

[29] Seaver, Nick. "Algorithms as Culture: Some Tactics for the Ethnography of Algorithmic Systems."
Big Data & Society 4(2), 2017. https://doi.org/10.1177/2053951717738104.

[30] Seaver, Nick. *Computing Taste: Algorithms and the Makers of Music Recommendation*. Chicago: University of Chicago Press, 2022

[31] Suchman, Lucy. *Plans and Situated Actions: The Problem of Human-Machine Communication*. Cambridge: Cambridge University Press, 1987

[32] Suchman Lucy. Human-Machine Reconfigurations: Plans and Situated Actions Cambridge, UK: Cambridge Univ. Press. 2007.

[33] Tsing, A. “More than Human Sociality: A Call for Critical Description.” In Anthropology and Nature, edited by Kirtsten Hastrup, 27-42. London: Routledge, 2013.

[34] Turner, Victor. *The Ritual Process: Structure and Anti-Structure*. Chicago: Aldine Publishing Company, 1969