1st Edition

InnoComp 2025 — Innovative Perspectives on Computational Intelligence and Data Science

Cluj-Napoca, Romania, October 22 – 24, 2025

BOOK OF ABSTRACTS

Organizers

Babeș-Bolyai University, Cluj-Napoca
https://www.ubbcluj.ro

Technical University of Cluj-Napoca
https://www.utcluj.ro

National University of Science and Technology Politehnica Bucharest
https://upb.ro

Partners

General Program

1st Edition

Innovative Perspectives on Computational Intelligence and Data Science
InnoComp 2025

BOOK OF ABSTRACTS

Editors

Camelia CHIRA, Oliviu MATEI, Florin POP, Petrică POP-SITAR

Cluj-Napoca • Romania • 22-24 October 2025

Table of Contents

A word from the Conference Chairs

InnoComp is an international conference that brings together researchers, practitioners, and thought leaders from academia, industry, and government to explore cutting-edge advancements and emerging challenges in the dynamic fields of computational intelligence and data science. The conference covers a spectrum of topics, including artificial intelligence, intelligent systems and data science, ensuring a holistic view of the rapidly evolving technological landscape. Attendees will be able to engage with leading experts through keynote speeches, panel discussions, hands-on workshops, and poster presentations.

The InnoComp conference focuses on cutting-edge advancements and emerging challenges in the dynamic fields of computational intelligence and data science. The conference covers a spectrum of topics, including artificial intelligence, intelligent systems and data science, ensuring a holistic view of the rapidly evolving technological landscape.

We are confident that this first edition of the conference will open new perspectives of collaboration and will create a proper environment to organize the next editions.

The conference volume will be published by Springer in *Communications in Computer and Information Science* (CCIS) series. It includes the papers accepted for presentation at InnoComp 2025 - the 1st International Conference on Innovative Perspectives on Computational Intelligence and Data Science, held in Cluj-Napoca, Romania during October 22-24, 2025.

InnoComp 2025 received more than 120 submissions. After a rigorous peer review process, the International Program Committee selected 49 full papers which are published in this conference proceedings. The submitted papers were carefully reviewed to keep a high-quality level of the conference, and we would like to thank the Program Committee and to the additional reviewers for their hard work during the reviewing process. The papers have been organized on the following topical sections: (i) Computational Intelligence, (ii) Data Science, (iii) Intelligent Systems, and (iv) Applications.

InnoComp 2025 featured outstanding keynote talks by three distinguished invited speakers: Prof. Jose Antonio Barata de Oliveira – NOVA University of Lisbon (Portugal), Prof. Aniello Castiglione – University of Salerno (Italy) and Prof. Andrei Paun - University of Bucharest (Romania).

The book of abstracts offers to all participants the possibility to have a general view on the accepted papers and to focus on main research contribution included in each paper.

Organizing this first edition of the international conference is a collective effort. We are honored to thank all the members of the organization committee and program committee, as

well as all the reviewers, for their hard work in finalizing the reviews on time, and the authors for submitting their papers. This wonderful community of researchers contributes to the success of this event. We also express our gratitude and thank all members of the local organizing committee for their valuable contribution to this successful event.

The editors would like to thank to Ronan Nugent (Springer Heidelberg, Editorial Director) for the editorial assistance and excellent cooperative collaboration to produce this valuable scientific work. We express our thanks to all Springer representatives who help us during the production of this conference. We appreciate the support offered by the Springer Meteor system team to handle the paper submission, review process, and communications with authors and reviewers. We thank them for this important support.

We hope you will enjoy the conference together with us and we hope to establish new connections and collaborations during this event.

October 2025

Camelia Chira
Oliviu Matei
Florin Pop
Petrică Pop-Sitar

Chairs and Committees

General Chairs

Camelia Chira, *Babeş-Bolyai University, Romania*
Oliviu Matei, *Holisun SRL, Romania*
Florin Pop, *National University of Science and Technology Politehnica Bucharest, Romania*
Petrică Pop-Sitar, *Technical University of Cluj-Napoca, Romania*

Program Committee

Lenuța Alboaie, *Alexandru Ioan Cuza University, Romania*
Apostolos Ampatzoglou, *University of Macedonia, Greece*
Anca Andreica, *Babeș-Bolyai University, Romania*
Costin Badica, *University of Craiova, Romania*
Ovidiu Bagdasar, *University of Derby, UK*
Constantin Bala Zamfirescu, *Lucian Blaga University of Sibiu, Romania*
Jose Antonio Barata De Oliveira, *University Nova De Lisboa, Portugal*
Ana Bazzan, *Federal University of Rio Grande Do Sul, Brazil*
Ion Bica, *Military Technical Academy, Romania*
Mihaela Breabăn, *Alexandru Ioan Cuza University, Romania*
Alexandros Chatzigeorgiou, *University of Macedonia, Greece*
Camelia Chira, *Babeș-Bolyai University, Romania*
Reggie Davidrajuh, *University of Stavanger, Norway*
Marc Demange, *Royal Melbourne Institute of Technology, Australia*
Liviu Dinu, *University of Bucharest, Romania*
Laura Dioșan, *Babeș-Bolyai University, Romania*
Ciprian Dobre, *National University of Science and Technology Politehnica Bucharest, Romania*
Enol García González, *University of Oviedo, Spain*
Dorian Gorgan, *Technical University of Cluj-Napoca, Romania*
Alvaro Herrero, *University of Burgos, Spain*
Vasile Manta, *Technical University Gheorghe Asachi, Romania*
Oliviu Matei, *Holisun SRL, Romania*
Daniel Mican, *Babeș-Bolyai University, Romania*
Mihai Micea, *Polytechnic University of Timișoara, Romania*
Sanaz Nikghadam Hojjati, *University Nova de Lisboa, Portugal*
Constantin Orasan, *University of Surrey, UK*
Michele Pagano, *University of Pisa, Italy*
Vincenzo Pallotta, *University of Applied Science and Arts, Switzerland*
Andrei Păun, *University of Bucharest, Romania*
Dana Petcu, *West University of Timișoara, Romania*

Florin Pop, *National University of Science and Technology Politehnica Bucharest, Romania*
Petrică Pop-Sitar, *Technical University of Cluj-Napoca, Romania*
Radu Emil Precup, *Polytechnic University of Timişoara, Romania*
Hector Quintian, *University of a Coruña, Spain*
Dimitrios Raptis, *Aalborg University, Denmark*
Günter Rudolph, *Technical University Dortmund, Germany*
Stefano Secci, *Conservatoire National des Arts Et Métiers, France*
Gabriel Semanišin, *Pavol Jozef Šafárik University, Slovakia*
Dragan Simić, *University of Novi Sad, Serbia*
Andrei Sleptchenko, *Khalifa University Abu Dhabi, UAE*
Ioannis Sorokos, *Fraunhofer IESE, Germany*
Giovanni Stilo, *University of L'Aquila, Italy*
Nikolai Velikov, *Nikola Vaptsarov Naval Academy, Bulgaria*
José Ramón Villar, *University of Oviedo, Spain*
Michał Woźniak, *Wroclaw University of Science and Technology, Poland*
Daniela Zaharie, *West University of Timişoara, Romania*

Additional Reviewers

Diana Borza
Alina Călin
Cerasela Crişan
Cristina Feier
Bela Genge
Barna Iantovics
Elena Nechita
Mara Hajdu-Măcelaru
Anca Mărginean
Tudor Mihoc
Gabriela Moise
Simona Motogna
Mădălina Răschip
Mihai Suciu
Adrian Petrovan

Organizing Committee (Babeş-Bolyai University, Romania)

Camelia Chira
Anca Andreica
Laura Dioşan
Simona Motogna
Cristiana Moroz-Dubenco
Mădălina Dicu
Raluca Chiş
Alexandru Manole
Oskar Picus
Ciprian Stupinean

1st Edition

Innovative Perspectives on Computational Intelligence and Data Science
InnoComp 2025

Plenary Speakers

Frontiers in the Quantum Computing Era

Prof. Aniello CASTIGLIONE
University of Salerno, Italy

Abstract **-** A new era of computing is on the horizon, as quantum computation introduces opportunities in optimization, drug discovery, and secure communication, while posing challenges in cryptographic security, error correction, and system scalability. Meanwhile, as quantum computing advances, many widely used cryptographic schemes, such as RSA and elliptic curve cryptography, are expected to become vulnerable to quantum attacks, particularly through Shor's algorithm. To address this looming threat, post-quantum cryptography (PQC) has emerged as a field dedicated to designing cryptographic algorithms that can withstand both classical and quantum adversaries, ensuring long-term security in the quantum era. Practical applications of post-quantum cryptography (PQC) includes the protection of critical IoT devices, ensuring the integrity and authenticity of machine learning models, and enhancing resilience against ransomware attacks.

For IoT devices, adapting PQC algorithms to low-power and low-resource environments is essential. This adaptation involves developing lightweight cryptographic solutions that maintain security without compromising device performance. Furthermore, one of the major challenges in Federated Learning (FL) is the security risks arising from model exchanges between clients and the central server or among clients themselves. Thus, integrating PQC into Federated Learning (FL) workflows can safeguard model integrity by preventing unauthorized alterations and ensuring data authenticity of the decentralized knowledge discovery. Finally, the cryptographic schemes used by ransomware in their operations, such as data encryption, exfiltration, and Command & Control communication, are also potentially vulnerable to quantum attackers. All these examples create a pressing need to transition to cryptographic primitives that are resistant to quantum attacks, ensuring the long-term security of sensitive data and mitigating the risk posed by future quantum-enabled threats.

Biography **-** Prof. Aniello Castiglione is an Associate Professor in the Department of Management & Innovation Systems at the University of Salerno, Italy. He earned his PhD in Computer Science from the University of Salerno in 2007. His principal research areas include information forensics, digital forensics, security and privacy (especially in cloud and network environments), communication networks, applied cryptography, and sustainable computing. Dr. Castiglione has published broadly in international journals and conferences (~190+ papers) and has held multiple editorial and organizational roles for conferences and journals in his fields of expertise.

New research directions examples from ICIA: using AI in Biomedicine and other fields

Prof. Andrei PĂUN
University of Bucharest, Romania

Abstract - In this presentation, we highlight emerging and evolving research directions within the Institute for Computational Intelligence and Applications (ICIA) over the past year. These interdisciplinary efforts span a diverse array of application domains, including medicine, neuromorphic computing, bio-computing, and the detection of deep fakes. We will discuss both novel and ongoing initiatives that exemplify ICIA's commitment to advancing foundational theory while tackling pressing real-world challenges. Special attention will be given to how computational paradigms inspired by biology are shaping the future of intelligent systems, and how these innovations are being translated into impactful applications in healthcare and digital security.

Biography - Andrei Păun is professor of Computer Science at University of Bucharest and Director of the Research Institute for Artificial Intelligence "Mihai Drăgănescu" (ICIA) (since 2024). Andrei's education and research forming took place on two continents (Europe and North America) and in multidisciplinary research groups from four universities (University of Bucharest: Math and Computer Science, University of Western Ontario: Computer Science and Biology, Rovira y Virgili University: Computer Science and Linguistics, Louisiana Tech University: Computer Science and Biology). His main interests are AI, Biocomputing, Bioinformatics, as well as finite automata and other related areas. Andrei organised/co-organised conferences in automata theory, bioinformatics/Systems Biology both in North America and Europe, the most recent being CMSB 2022. Aside from the deep results in the interdisciplinary areas of Computer Science, Medicine and Biology known as Biocomputing, Bioinformatics and Systems Biology Dr. Paun showed a strong versatility to work with various research groups from many disciplines, especially Biology and Medicine.

Smart Manufacturing: Current Challenges on Using Computational Intelligence Jose Antonio BARATA De OLIVEIRA
Universidade NOVA de Lisboa, Portugal

Abstract **-** Smart Manufacturing represents the convergence of digital technologies, intelligent control, and adaptive systems engineering. While the fourth industrial revolution has provided unprecedented digital connectivity, the next frontier involves the integration of computational intelligence to enable autonomy, resilience, and self-organization. This presentation discusses the current challenges in embedding intelligence within complex manufacturing systems. It argues that Smart Manufacturing must be understood as a complex adaptive system, where emergent behaviors arise from interactions between distributed entities. The presentation further discusses how computational intelligence—ranging from biologically inspired algorithms to multi-agent architectures—can address adaptation, learning, and evolution in manufacturing ecosystems. Contributions of the author in the fields of holonic and agent-based control, self-organization, and biologicalisation of manufacturing are also presented as practical exemplars of this long road toward intelligent factories.

Biography **-** José António Barata de Oliveira is Full Professor in the Department of Electrical and Computing Engineering (DEEC) at NOVA School of Science and Technology (NOVA-FCT), Universidade NOVA de Lisboa, Portugal. He is a senior researcher at the Centre for Technology and Systems (CTS/UNINOVA), where he leads the research group RICS — Robotics and Industrial Complex Systems. His research interests include robotics, autonomous systems (e.g. UAVs, service robots), cyber-physical systems, intelligent manufacturing / Industry 4.0, perception and navigation, and applications such as smart agriculture. Professor Barata's academic formation includes a PhD in Electrical and Computer Engineering from NOVA University of Lisbon (2004), a Master's in Computer Science from the same institution (1995), and earlier degrees in Computer Science and Electrical Engineering / Digital Systems from NOVA and ISEL. He also holds the honorary title Doctor Honoris Causa awarded by the Technical University of Cluj-Napoca in 2023.

1st Edition

Innovative Perspectives on Computational Intelligence and Data Science
InnoComp 2025

Abstracts

Computational Intelligence

Analysis of Customer Journey Video Data Using Eye-Tracking and Multimodal AI

Sami Heikkinen, Jaani Väisänen, Hannu Kaikonen, Sami Makkula, Petteri Markkanen
LAB University of Applied Sciences, Lahti, Finland
sami.heikkinen@lab.fi

Abstract – This study introduces a novel methodological approach for analyzing customer journey data using eye-tracking technology and multimodal AI. Customer journey research faces significant limitations due to reliance on resource-intensive qualitative methods that are difficult to scale. We address this gap by employing vision language models (VLMs) to automate the interpretation of eye-tracking data, eliminating the manual coding bottleneck while enabling analysis of larger and more diverse datasets. Our research evaluates five locally-deployable VLMs to determine the optimal balance between semantic accuracy and computational efficiency. Using data collected with Tobii eye-tracking glasses, we developed an analytical pipeline that synchronizes gaze location data with verbal think-aloud protocols to create a comprehensive multimodal dataset. Results demonstrate that Gemma3 (4B parameters) achieved 100% semantic accuracy on our test set while maintaining reasonable processing efficiency (43.26 seconds per image). When validated against human coding across the complete dataset, the model achieved a 74.2% recall rate. The integration of eye-tracking and verbal data revealed distinctive attention patterns including "navigational uncertainty," "confirmatory scanning," and "socially-mediated attention" throughout the customer journey. Our approach provides objective behavioral evidence of visual attention that complements traditional self-reported measures, enabling more comprehensive touchpoint analysis while aligning with event-driven perspectives from process mining research. This methodology offers promising applications for service design by identifying discrepancies between reported and actual customer attention patterns and providing a foundation for developing automated behavioral indicators to detect moments of customer confusion, decision-making, or confirmation throughout service journeys.
Keywords – Customer journey analysis, Eye-tracking, Vision language models, Service design, Customer experience.

Efficient Face Recognition with ResNet18

Ioan Chiţu[1,2], Honorius Gâlmeanu[1,2], and Alexandru Drîmbărean[3]
[1]Department of Mathematics and Informatics, Transilvania University of Braşov, Romania
[2]FotoNation SRL, Braşov, Romania
[3]FotoNation LTD, Ireland
chituioan17@gmail.com, galmeanu@unitbv.ro, alexandru.drimbarean@fotonation.com

Abstract – The domain of face recognition has shown a dramatic development in the recent years. Most impressive results were encountered using deep convolutional networks. Driven by reasons of efficiency, we approached the problem using a simple architecture, namely a ResNet18 network with 11 million parameters. By using the Additive Margin Soft-max loss function we show how the performance of a plain architecture with no changes can be improved close to the one of state-of-the-art models with an order of magnitude more parameters. We

obtain a train accuracy of 96% and a validation accuracy of 87% and a ROC area close to 0.98 on the LFW dataset. We show that even such a simple architecture is appropriate for face recognition.
Keywords – Deep CNN, image processing, face recognition, ResNet18.

Hierarchical Multi-Task Siamese Networks for Vehicle Model Classification

Alexandru Manole, Laura Dioșan
Department of Computer Science, Babes-Bolyai University, Mihai Kogalniceanu nr.1, Cluj-Napoca, Romania
{alexandru.manole, laura.diosan}@ubbcluj.ro

Abstract – This work investigates a hierarchical Siamese multi-task learning approach for vehicle model similarity, using the StanfordCars dataset as a benchmark. The main contribution is the combined learning objective, which aims to jointly learn to compute the similarity between two, or more hierarchic concepts (i.e., car make comparison and the fine-grained car model similarity), a novel technique, especially for vehicle model classification. Two multi-task architectures are presented and evaluated: parallel and cascaded, which incorporate auxiliary objectives such as car make and car type similarity to enhance the main task of model-level similarity estimation. The experiments span nine established classifiers and examine both single-task and multi-task configurations. The results demonstrate that multi-task learning improves performance by up to 8.8% over single-task baselines. Notably, four of the top five configurations leverage a multi-task setup, with the Parallel EfficientNet model, and make similarity objective achieving the highest accuracy (86.9%). Incorporating car make as an auxiliary objective consistently boosts model similarity accuracy by 1–2%, and parallel architectures outperform cascade in most scenarios. Given the complexity of vehicle make and model classification (VMMC), a domain with hundreds of thousands of class variants and rapid model turnover, the findings support multi-task learning as a scalable solution. This is particularly relevant for intelligent transportation systems, where accurate recognition of vehicle types is essential for applications in traffic monitoring, driver assistance, and smart city applications.
Keywords – Multi-Task Learning, Siamese Networks, Vehicle Model Similarity.

A comparison of ranking methods for feature selection

Grace Villacrés[1], Marius Marinescu[1], Luca Martino[2], Óscar Barquero[1]
[1]Universidad Rey Juan Carlos, Fuenlabrada, Spain.
[2]Università di Catania, Italy.
grace.villacres@urjc.es

Abstract – Feature selection, also known as variable selection, plays a critical role in statistics, signal processing, and machine learning by enhancing model interpretability, identifying relevant variables for specific tasks, and enabling effective dimensionality reduction. Central to this process is the accurate ranking of variables. In this study, we conduct a systematic and fair comparison of several wrapper-based feature selection methods to evaluate their performance. To ensure a controlled and insightful analysis, we generate synthetic datasets that incorporate variable correlations, different distributions, and interdependencies. Our findings reveal that two specific wrapper methods consistently outperform others, offering practical guidance for both researchers and practitioners. Notably, the widely discussed leave-one-covariate-out (LOCO) method—closely related to Shapley value-based approaches—performs the worst in our evaluations. In contrast, the backward elimination method, which constructs a variable ranking by iteratively removing the least important features, achieves the highest

performance. These results underscore the importance of method selection in feature ranking and provide actionable insights for future applications.
Keywords – Feature selection, Ranking method, Wrapper methods.

Automated Model Interpretation, Pattern Analysis and Insights in Explainable AI

Iulian Popa[1], David-Ioan Stoica[2], Alexandra Băicoianu[2]
[1]Faculty of Mathematics and Informatics, Transilvania University of Braşov, România
[2]Department of Mathematics and Informatics, Transilvania University of Braşov, România
david.stoica@student.unitbv.ro, {iulian.popa, a.baicoianu}@unitbv.ro

Abstract – Deep learning has revolutionized artificial intelligence by enabling highly accurate models across diverse domains such as computer vision and autonomous driving. However, the inherent complexity of deep neural networks often results in *black-box* models, whose decision-making processes lack transparency and interpretability. Explainable Artificial Intelligence addresses this challenge by developing techniques that make model predictions more understandable to humans. Among these, Class Activation Mapping (CAM) and its variants: Grad-CAM, Grad-CAM++ and Score-CAM, have become pivotal in providing visual explanations for convolutional neural networks. This paper presents a systematic comparison of several prominent CAM-based explanation methods applied to the AlexNet architecture on a standardized Dogs vs. Cats image dataset. We evaluate these methods using quantitative metrics including Intersection over Union, Dice coefficient and Deletion and Insertion tests, aiming to reveal their respective strengths, limitations and practical utility. These findings offer valuable insights to guide practitioners in selecting appropriate explanation techniques. Specifically, we show that combining CAM-based techniques enables a deeper understanding of pixel importance by identifying critical regions composed of both object and background pixels and by guiding data preprocessing promotes more stable training and improves the model's focus. Furthermore, we demonstrate the potential for constructing robust feature attribution strategies grounded in layer-wise analysis and assess the reliability of various XAI algorithms by evaluating their behavior across multiple threshold levels.
Keywords – Explainable AI, CAM techniques, Saliency maps.

Pairwise 3D Fragment Matching Classification with Graph Neural Networks

Raluca Chis, Camelia Chira
Babeş-Bolyai University, Cluj-Napoca, Romania
raluca.chis@ubbcluj.ro, camelia.chira@ubbcluj.ro

Abstract – Automatic object reassembly is an important task in archaeology, robotics and medicine. The creation of a tool for finding fragments that complement each other and assembling them into a more or less complete object would facilitate the work of restorers. Recent works focus on predicting the pose of a set of fragments relative to each other or on estimating the matching of different surfaces with mathematical approaches. With this article, we propose a deep-learning method for pairwise matching classification of point clouds. The research methodology is based on the use of Siamese neural networks combined with graph convolutions. Pairs of matching and non-matching point clouds have been created using the fragmented objects provided in the Breaking Bad dataset. These objects are represented as point clouds and are divided into multiple pieces. Our best experiment reveals an accuracy of 82.7% and an F1 score of 81.09%, which demonstrates the power of the implemented method.
Keywords – Point cloud matching, Graph networks, Breaking Bad.

Music Emotion Recognition with Deep Learning Techniques

Denis-Angel Moldovan, Claudia-Ioana Coste
Babeș-Bolyai University, Cluj-Napoca, Romania
denis.angel.moldovan@stud.ubbcluj.ro, claudia.coste@ubbcluj.ro

Abstract – Music is considered an *universal language*, a way to reduce language and cultural barriers between people from around the globe. The present paper proposes to dive into the subjective domain of emotion recognition, aiming to decipher the relationship between musical elements and human emotions. In the present paper, deep learning methods are leveraged to classify the mood induced by a piece of music. For feature modelling, multiple attributes are used. Spotify features are employed, such as rhythm, tone, danceability, energy, loudness, valence, speechiness, and others. Other features, such as Mel-Frequency Cepstral Coefficients, Spectral Centroid, Chroma Energy Normalized Statistics, and the Mel Spectrogram, are concatenated into an image. The conducted experiments employed deep learning methods, including a ResNet18 model, which had not been used in this context before. The best-performing model, with 86% accuracy, proves to be a hybrid one, based on ResNet18 architecture and enriched with two Bidirectional Long-Short Term Memory layers, outperforming by 6%-11% other deep learning architectures that the present paper experimented with. The results are rather similar to other studies from the literature, even though the approaches and the datasets are different.
Keywords – Music emotion recognition, deep learning, information retrieval.

UI Component Detection in Graphical Interfaces: A YOLO-Based Comparative Analysis

Madalina Dicu
Faculty of Mathematics and Computer Science, Babeș-Bolyai University, Cluj-Napoca 400084, Romania
madalina.dicu@ubbcluj.ro

Abstract – In recent years, detecting visual components in graphical user interfaces (GUIs) has become increasingly relevant for modern applications that rely on automatic UI understanding. Unlike natural images, UI layouts present unique challenges, such as small, densely packed elements and high visual variability, which can affect standard object detectors. To explore this problem, we conduct a comparative analysis of six recent YOLO (You Only Look Once) versions, ranging from YOLOv5 to YOLOv12, for the task of detecting graphical components in user interfaces. We target three major platforms (mobile, web, desktop) and use representative datasets: VINS (mobile), GENGUI (web), and UICVD (desktop). All annotations were standardized to include the same three core UI classes: `Text`, `Button`, and `Icon`. Each YOLO model was trained and evaluated separately on each dataset, using the lightweight `nano(n)`, `tiny(t)`, and `small(s)` variants to analyze the trade-off between accuracy and efficiency. The results show noticeable performance differences across versions and interface types. YOLOv9s and YOLO11s performed best in complex scenarios, while YOLOv5s excelled on mobile data. `Button` was the easiest class to detect, `Text` remained stable, and `Icon` proved the most challenging. Several models, including YOLOv9s and YOLO11s, achieved mean Average Precisions (mAP) above 93% across desktop, web, and mobile interfaces, confirming that detection quality depends strongly on both the architecture and the interface type.
Keywords – Object detection, UI detection, YOLO, mobile, web, desktop.

Automated Classification of Romanian Mammography Reports

Cristiana Moroz-Dubenco, Anca Andreica
Faculty of Mathematics and Computer Science, Babeș-Bolyai University, Cluj-Napoca, Romania
cristiana.moroz@ubbcluj.ro, anca.andreica@ubbcluj.ro

Abstract – This study presents an automated system for classifying breast cancer diagnosis and BI-RADS scores based on free-text mammography reports written in Romanian. Utilizing a newly collected and publicly available Romanian Dense Breast Mammography Collection (RDBMC) dataset, we define two classification tasks: (1) assessing the patient's overall condition as healthy, benign, or malignant, and (2) predicting the severity of findings using the BI-RADS scoring system. To address the scarcity of research on Romanian medical reports, we experiment with both the original Romanian texts and their English translations. We evaluate four classifiers (Random Forest, Logistic Regression, Decision Trees, and Naive Bayes) combined with two text embedding methods (TF-IDF and Latent Semantic Indexing), incorporating patient age as an additional feature. Our results show that the Random Forest classifier with TF-IDF embeddings achieves the best performance for Romanian reports, with an accuracy of 80% in diagnosis classification and 73% in BI-RADS scoring, demonstrating robustness and potential clinical applicability. Although English translations yield slightly higher scores, differences are small, indicating that the system is effective across languages.
Keywords – Mammography Reports, Medical Text Classification, Romanian Language, Breast Cancer Diagnosis, BI-RADS Scoring.

Comparative Analysis of Activation Functions in Neural Networks for Computer Vision Applications

Andrei Sebastian Iaz, Octavia-Maria Bolojan, Robert Valentin Neagu
University of Oradea, Oradea, Romania
iaz.andreisebastian@student.uoradea.ro, obolojan@uoradea.ro, neagu.ro, bertvalentin@student.uoradea.ro

Abstract – Activation functions play a pivotal role in the performance and convergence of deep neural networks (DNNs) as well as convolutional neural networks (CNNs). In this study, a comprehensive comparative analysis is conducted on fourteen activation functions, including both classical functions, for example, ReLU, Sigmoid, Tanh and recent proposals such as Si-LU, Mish, Serf, Nipuna. We apply these functions to a DNN architecture trained on the MNIST dataset, and then on a CNN trained on both MNIST and CIFAR-10. Results include accuracy, training time, inference speed, and the cumulative loss area (computed via numerical integration). The results indicate that while some functions deliver high accuracy and an efficient convergence, others result in statistically significant decreases in accuracy and prolonged convergence times. This paper outlines the methodology, provides the mathematical framework, and presents and discusses the findings.
Keywords – Accuracy, Activation Function, Convergence, Deep Learning, Digit Recognition, MNIST, Neural Networks.

Embedding-Efficient Brain-to-Image Reconstruction Using Diffusion Models

Ioana Gabor
Department of Computer Science, Faculty of Mathematics and Computer Science, Babes-Bolyai University, Cluj-Napoca, Romania
ioana.gabor.cj@gmail.com

Abstract – Visual brain decoding is the process of reconstructing the images a person sees by analyzing their brain activity, as measured using functional MRI, electroencephalography, or

magnetoencephalography. Recent advances in generative AI, particularly diffusion models, have significantly improved the semantic quality of the reconstructed images. In this work, we investigate the potential for improving the computational efficiency of brain-to-image models by focusing on a reduced number of informative embeddings. Specifically, we explore whether high-quality image reconstructions can be achieved by predicting a small, selected subset of CLIP-derived embeddings. We perform evaluations using the Natural Scenes Dataset, applying both a simple single-subject ridge regression model and a multi-subject federated learning framework adapted from a recent approach. The results demonstrate that embedding-efficient decoding can achieve competitive performance while substantially reducing model size, although with a decrease in quality on certain metrics. Code available at: https://github.com/IoanaGabor/efficient-vbd.
Keywords – Brain–computer interfaces, functional magnetic resonance imaging (fMRI), image reconstruction, diffusion models, neural decoding, CLIP embeddings, federated learning, ridge regression.

Fine-tuned Vision Transformers for Mammographic Breast Cancer Classification: Evaluating the Impact of Preprocessing Techniques

Bianca Iacob, Laura Dioșan
Faculty of Mathematics and Computer Science, University Babes Bolyai, Cluj-Napoca, Romania
lixandrubianca01@gmail.com, laura.diosan@ubbcluj.ro

Abstract – This study examines the comparative performance of Vision Transformers in classifying breast cancer images under various pre-processing conditions. The analysis was conducted on raw breast cancer mammograms, as well as images enhanced using Gamma correction and Contrast Limited Adaptive Histogram Equalization. Specifically, five dataset configurations were considered: (1) original images, (2) Gamma-corrected images only, (3) original images combined with Gamma-corrected versions, (4) Contrast Limited Adaptive Histogram Equalization-enhanced images only, (5) original images combined with Contrast Limited Adaptive Histogram Equalization-enhanced versions. Additionally, the analysis was stratified by lesion types and mammographic view, comparing model performance for calcification and mass lesions on Mediolateral Oblique and Craniocaudal projections. The results reveal distinct performance patterns in the vision transformer model across these preprocessing scenarios and imaging perspectives, highlighting the impact of image enhancement techniques and view types on model performance. This study provides valuable insights into the effectiveness of pre-processing methods in optimising the performance of advanced neural network architectures for breast cancer image classification. The model demonstrates strong performance across both datasets, achieving its highest results on MIAS dataset with an accuracy, sensitivity, and specificity of 0.99 and on DDSM dataset with an accuracy of 0.95, sensitivity of 0.96 and specificity of 0.95.
Keywords – ViT, CLAHE, Gamma correction, breast cancer, mammograms.

Data Science and Computational Intelligence

Cultivating Sustainable Data: Challenges and Directions in Digital Agriculture

Daniela Delinschi[1,3], Rudolf Erdei[1,3], Emil Marian Pasca[1,3], Iulia Baraian[1,3], Andrei Curos[2]
[1]Technical University of Cluj-Napoca, Romania
[2]University of Petrosani, Romania
[3]European University of Technology, European Union
daniela.delinschi@campus.utcluj.ro, rudolf.erdei@ieec.utcluj.ro, emil.pasca@campus.utcluj.ro, iulia.baraian@campus.utcluj.ro, andreicuros@upet.ro

Abstract – This paper addresses data sustainability in smart agriculture, examining challenges and best practices throughout the agricultural data lifecycle. It differentiates between efficiency, sustainability, and resilience in data systems while analyzing key sustainability challenges including infrastructure limitations, data quality issues, and ownership concerns. The research proposes and details a conceptual framework for quantifying data sustainability, based on dimensions such as data relevance, utility, criticality, age, and cost. A Romanian agriculture illustrative application illustrates sustainability challenges across different farm scales with targeted improvement strategies. The findings contribute to understanding how sustainable data management can enhance agricultural practices while balancing technical, environmental, economic, and social dimensions for long-term viability.
Keywords – Data Sustainability, Agricultural Data Lifecycle, Smart Agriculture, Data Management, Precision Agriculture, Digital Agriculture.

Towards Self-learning AI-based Quality Rules for Statistical Process Control

Marcel Dechert[1], Fulya Horozal[1], Sebastian Scholze[1], Marcel Wabo[2], Ratan Kotipalli[2]
[1]Institut für angewandte Systemtechnik Bremen GmbH, Bremen, Germany
[2]Continental Automotive Technologies GmbH, Babenhausen, Germany
dechert@atb-bremen.de, horozal@atb-bremen.de, scholze@atb-bremen.de, marcel.wabo@continental-corporation.com, raja.venkata.ratan.kotipalli@continental-corporation.com

Abstract – Amid the immense complexity of modern Industry 4.0 manufacturing, traditional Statistical Process Control (SPC) approaches, which are reliant on manually defined thresholds and expertise-driven control charts, are proving increasingly insufficient. The vast amount of data, continual evolution of products and processes and the necessity for timely, meaningful alerts have exposed key limitations in existing methods, including scalability, adaptability and transparency. This paper introduces a novel automated, AI-driven methodology called Self-learning AI-based Q-Rules (SLAQ) to overcome these challenges. The SLAQ methodology combines optimization techniques with reinforcement learning to dynamically generate and refine quality rules (Q-Rules) for anomaly detection in production data. Initial Q-Rules are derived from historical data via optimization algorithms. These rules are subsequently refined through reinforcement learning using human feedback to enhance detection performance and adaptability to address changing process conditions.
Keywords – Statistical process control, data science, AI, reinforcement learning.

Modeling Regional Smart Specialization via Value Chains: A Stochastic Approach Using Markov Chains for Smart Development Planning

Laura Andreica, Mara Diana Hajdu-Macelaru, Mario Viman, Mihai Crisan, Ioan Alin Danci
Technical University of Cluj-Napoca, Romania
laura.andreica@campus.utcluj.ro

Abstract – This paper presents an innovative, data-driven framework for determining the Economic Smart Specialization (ESS) of local communities through the use of discrete-time Markov chains. By synthesizing three critical layers of information – intrinsic industrial specialization derived from NACE-based revenue data, inter-industry supply chain flows – the proposed model captures the complex interplay between economic structure and geographic proximity. Specialization is modeled as a stochastic process, enabling simulation of how sectoral strengths evolve and diffuse across regions. The framework yields a normalized ESS matrix that provides a probabilistic index of specialization at the locality level, informed by both upstream and downstream industrial relationships. Through Principal Component Analysis (PCA) and relative standard deviation metrics, the model quantifies the diversity and complexity of economic specialization under multiple transition scenarios. Applied to more than 3000 localities in Romania, the methodology demonstrates high scalability and adaptability, offering policy makers a powerful tool to design targeted, resilient, and context-sensitive smart specialization strategies. This research bridges regional economic modeling and machine learning, contributing a reproducible, interpretable, and policy-relevant approach to enhancing territorial innovation potential in line with the principles of the Smart Specialization Strategy (S3).

Keywords – Markov Chains, Stochastic Algorithms, Economic Graph Modeling, Matrix Factorization, Specialization Inference, Data-Driven, Policy Modeling.

Advancing Natural Language Querying over Invoice Data: Experimental Insights on Vector and Graph RAG Approaches

Teodora-Cristiana Nemtoc, Ana-Maria Ghiran
Faculty of Economics and Business Administration, Babeș-Bolyai University, Cluj-Napoca, Romania
cristiana.nemtoc@econ.ubbcluj.ro, anamaria.ghiran@econ.ubbcluj.ro

Abstract – The expanding integration of AI techniques into daily organizational processes, has transformed the way business-related questions are addressed, significantly reducing the need for technical expertise. In this paper, we evaluate the extent to which responses to stakeholder-relevant questions can be addressed using Retrieval-Augmented Generation (RAG) and its variant, GraphRAG, for information extraction from financial documents such as invoices. This study assesses the performance of these approaches at varying levels of complexity and question categories, identifying how factors such as structure, complexity, and the processes involved in the query pipeline impact their effectiveness. Particular attention is given to how Graph-RAG broadens the scope of answerable questions by structuring documents through graph representations. Guidelines are also included on how each procedure can be integrated into financial workflows, highlighting their respective strengths and limitations compared to a SPARQL benchmark approach.

Keywords – Invoice processing, RAG, GraphRAG, Knowledge Graphs, Information Extraction.

Analysis of economic proximity contagion and regional level using eignevectors

Laura Andreica, Gabriel Andreica, Patrick Axinte, Ioan Alin Danci, Bogdan Vaduva
Technical University of Cluj-Napoca, Cluj-Napoca, Romania
laura.andreica@yahoo.com, gabrielandreica22@gmail.com, axintepatrick@gmail.com, alin.danci@gmail.com, bogdan.vaduva@ieec.utcluj.ro

Abstract – This paper introduces a novel spectral framework for analyzing the spatial dynamics of Smart Specialization (S3) diffusion across Romanian regions. While S3 has become a central pillar of EU cohesion policy, the mechanisms underlying inter-regional innovation contagion remain poorly quantified. We address this gap by integrating Principal Component Analysis (PCA) with eigenvalue–eigenvector decomposition to model economic proximity contagion among 42 counties and over 3,000 localities. A contagion matrix is constructed using population, Gross Value Added (GVA), and Geographic Distance to estimate influence potential and structural centrality within the national innovation ecosystem. Our results uncover non-obvious clusters of innovation diffusion, highlight Bucharest and Cluj-Napoca as central propagation hubs, and demonstrate how structural positioning and connectivity can outweigh raw economic mass in determining territorial impact. The methodology offers a scalable and reproducible approach for identifying regional leaders and followers in smart specialization, with clear implications for multilevel governance and targeted policy design. This work contributes to the operationalization of S3 contagion analysis through unsupervised learning and network-based modeling in transitional economies.
Keywords – PCA, Eigenvectors, Machine Learning, Smart Specialization, Economic Contagion, Regional Innovation Systems.

Comparison of Multi Decision Criteria Analysis Methods in Molecular Dynamics of Tritium Interaction with Beryllium at Tokamak Fusion Reactors

Alper Pahsa
Havelsan Inc., Ankara 06510, Türkiye
Mustafa Kemal Mah. Şehit Öğretmen Şenay Aybüke Yalçın Cad. No:39
apahsa@havelsan.com.tr

Abstract – The understanding plasma-structural material interaction in Tokamak fusion reactors is tough. This topic has been studied extensively, including molecular dynamic simulations. Plasma material's interaction with high temperature fluctuations changes the fusion reactor wall's surface composition. This alteration may damage reactor components. This research aims to develop advanced materials that can survive harsh conditions, prevent degradation, and improve reactor performance. This methodology is essential to understanding fusion reactor power facility development. To study how plasma affects material surface crystal structure, an atom-by-atom simulation model of tritium sputtering on beryllium reactor wall structures is needed. Molecular dynamics Multi Decision Criteria Analysis (MCDA) simplifies operational parameter understanding. Implementing these methods improves molecular dynamics simulation. Using molecular dynamic modeling parameters for sputtering Tritium plasma on a Beryllium-based model surface, we evaluated many MCDA approaches.
Keywords – MCDA in Tokamak Fusion Reactors, MCDA at Molecular Dynamics, MCDA at Plasma Material Interaction.

Compact Real-Time Speech Enhancement for ONNX

Andrei Filip[1,2], Honorius Gâlmeanu[1,2], and Alexandru Drîmbărean[3]
[1]Department of Mathematics and Informatics, Transilvania University of Braşov, Romania
[2]FotoNation SRL, Braşov, Romania
[3]FotoNation LTD, Ireland
andreipilif@gmail.com, galmeanu@unitbv.ro, alexandru.drimbarean@fotonation.com

Abstract – Real-time speech enhancement is crucial for applications such as VoIP calls and hearing aids, which require low-latency noise suppression on resource-constrained devices. We present a causal speech enhancement model based on Facebook's Denoiser (a Demucs-inspired architecture), modified for deployment as an ONNX model. Key contributions include fixing the input to 2 seconds of audio at 16 kHz (32,000 samples), removing dynamic input length computation by padding to a constant length (32,085 samples), and trimming outputs accordingly during the training process. The model uses a convolutional encoder-decoder architecture with a gated LSTM bottleneck, trained on the CSTR VCTK Corpus+DEMAND (Valentini) dataset with data augmentation, and fine-tuned on the Microsoft Scalable Noisy Speech Dataset (MS-SNSD). The enhanced model achieves high speech quality, achieving results that match the state-of-the-art benchmarks for causal speech enhancement. The proposed model achieves PESQ 2.996 (vs. 3.07 for the original Denoiser), STOI 0.944 (vs. 0.95), CSIG 4.016, CBAK 3.453, and COVL 3.525 on the Valentini test set on par with state-of-the-art causal methods while remaining compatible with ONNX Opset 9. We also implemented a model based on a GRU bottleneck and observed a slight performance loss for offline processing, but better results during the online evaluation compared to the LSTM based model.

Keywords – Speech denoising, LSTM, GRU, U-Net, Deep CNN, ONNX.

Intelligent Systems

OpenBot Waffle: A 3D Printable Open-Source ROS2 Mobile Robot

Bogdan Dura[1], Radu Boncea[1], Sebastian Balmus[1], Catalin Nitu[1], Daniela Timisică[1,2]
[1]National Institute for Research and Development in Informatics - ICI Bucharest
[2]National University of Science and Technology POLITEHNICA Bucharest
bogdan.dura@ici.ro, radu.boncea@ici.ro, sebastian.balmus@ici.ro, catalin.nitu@ici.ro, daniela.timisica@ici.ro

Abstract – OpenBot Waffle is an open-source, ROS2-based mobile robotics platform designed for research, education, and rapid prototyping. Built around the Raspberry Pi 5, the platform integrates key sensors including the Slamtec A1 LiDAR, Intel RealSense D435i depth camera, and Adafruit BNO055 IMU, alongside high-performance Dynamixel XL430-W250-T motors. The robot's modular, 3D-printed chassis enables easy assembly and hardware customization. OpenBot Waffle supports autonomous navigation and mapping using the ROS2 Nav2 stack, with SLAM, localization (AMCL), and path planning capabilities. It also features seamless teleoperation via keyboard or joystick and can be fully simulated in Gazebo for virtual testing.
Keywords – Mobile Robotics, Robot Operating System 2 (ROS2), SLAM, Open-Source Robotics, Navigation.

Human-Centered AI in STEM Education

Mihaela Avramova
NVNA "N.Y. Vaptsarov", Dept. of Electronics, Varna, Bulgaria
m.avramova@nvna.eu

Abstract – The deployment of empathic artificial intelligence (AI) in higher education environments introduces complex ethical considerations that extend beyond traditional educational technology concerns. This work introduces the novel E3 Educational Framework (Empathy Recognition, Response, Reciprocity) - a theoretical model for implementing educational AI in STEM-focused educational environments. The E3 framework distinguishes the framework from adaptive systems through bidirectional emotional exchange while preserving student agency and academic integrity. The framework establishes three integrated layers: multimodal emotion detection achieving 92% accuracy for learning-specific emotional states, contextual interface adaptation maintaining academic rigor, and controlled emotional reciprocity within appropriate pedagogical boundaries. A comprehensive implementation guideline for learning management system integration, FERPA-compliant privacy protection, and novel evaluation metrics combining academic outcomes and emotional wellbeing measures is provided. Simulation testing across programming courses demonstrates the framework's effectiveness in reducing student anxiety while improving persistence rates without compromising academic standards.
Keywords – Empathic AI, Educational Technology, Adaptive Learning, Student Support, Learning Analytics, Human-Computer Interaction.

Bayesian Ensemble Learning for Adaptive Time Series Forecasting

Mateusz Panasiuk
AI Investments, Warsaw, Poland
mateusz.panasiuk@aiinvestments.pl

Abstract – We present a Bayesian approach to ensemble learning for time series forecasting, designed to address the challenge of model adaptation in non-stationary environments. Classical ensemble methods typically assume fixed model parameters and static combination rules, which limits their adaptability under temporal drift. Our method generalizes ensemble construction by framing both sub-model weight assignment and parameter adaptation as problems of probabilistic inference. Using Bayes' theorem, we derive an online update rule that balances new observations with prior knowledge, enabling coherent uncertainty quantification. Sub-model errors are transformed into ensemble weights via a parametrized function, with parameters governed by Gaussian priors and updated using a maximum a posteriori (MAP) approximation. This allows for fast, gradient-based updates without discarding accumulated experience. We evaluate the approach on a short-horizon directional forecasting task across 40 foreign exchange time series using eight neural network-based sub-models. The proposed method consistently outperforms static ensembles, with small but statistically significant improvements in accuracy and increase in profits in simulated setting. Beyond empirical findings, this work contributes a tractable and theoretically grounded framework for adaptive ensemble learning with Bayesian guarantees.
Keywords – Ensemble Learning , Forecasting, Bayesian method.

AugRoSent: Boosting Romanian Sentiment Analysis through Advanced Data Augmentation

Andra-Gabriela Ursa, Laura Dioșan
Babeș-Bolyai University, Cluj-Napoca, Romania
andra.ursa@stud.ubbcluj.ro, laura.diosan@ubbcluj.ro

Abstract – Sentiment analysis is a critical Natural Language Processing (NLP) task that categorizes text into positive, negative or neutral sentiments. While sentiment analysis has been extensively studied for English, research on Romanian sentiment analysis remains limited, primarily due to the scarcity of linguistic resources and specialized tools. To address this gap, AugRoSent (Augmentation for Romanian Sentiment Analysis) introduces a dataset augmentation framework designed to enhance the robustness and accuracy of Romanian sentiment analysis models. Three distinct augmentation techniques are applied within AugRoSent: Easy Data Augmentation (EDA), leveraging simple textual operations; text generation using a Romanian-specific open-source Large Language Model (OpenLLM-Ro); and fine-tuned GPT-2 models adapted to Romanian linguistic nuances. Each method's effectiveness is evaluated on a benchmark Romanian sentiment analysis dataset using the standard metric accuracy, alongside a series of qualitative metrics. Experimental results demonstrate that dataset augmentation improves model performance. These findings highlight the value of advanced language modeling approaches for resource-constrained languages and provide insights into selecting optimal augmentation strategies for Romanian NLP applications.
Keywords – Augmentation, Romanian Sentiment Analysis, Natural language processing (NLP), Easy Data Augmentation (EDA), GPT-based augmentation, Machine translation.

Immersive Robotics: Real-Time Control of Robotic Systems via ROS2 and Unity

Andrei Ionut Mocanu, Catalin Nitu
National Institute for Research and Development in Informatics ICI Bucharest, Romania
ionut.mocanu@ici.ro, catalin.nitu@ici.ro

Abstract – This paper explores the integration of Virtual Reality (VR), Unity, and Robot Operating System 2 (ROS2) to create an immersive and scalable control interface for robotic systems. As the complexity of robots increases, the basic simulation and control environments such as Gazebo often fail to deliver intuitive or complete representations of real-world challenges. In response, we propose an architecture that allows real-time, bi-directional communication between the user and a robotic arm – either physical or simulated – through a VR interface. Unity is employed as a visual and interactive platform, offering flexibility through community-developed plugins for both VR and ROS2. ROS2 provides robust, real-time communication capabilities with support for distributed systems, enhanced security, and industrial scalability. Together, they form a powerful ecosystem for developing human-robot interaction systems. Use cases in education and industrial robotics are presented to demonstrate the effectiveness of the proposed system. In educational settings, VR reduces costs and risks while enhancing engagement and skill acquisition. In industrial contexts, remote operation through VR improves safety, situational awareness, and control accuracy. This type of architecture confirms that integrating VR with ROS2 and Unity is not only feasible but highly beneficial for modern robotic applications, particularly within the scope of Industry 4.0. Future work includes adding AI-driven control, gesture recognition, and haptic feedback.
Keywords – Virtual reality, immersive technologies, robotic systems, Unity, ROS.

Intelligent Cyber-Social Ecosystem Industry 5.0

Neli Stoycheva
Nikola Vaptsarov Naval Academy, Department of Electronics, Varna, Bulgaria
n.stoycheva@naval-acad.bg

Abstract – This article explores the emerging paradigm of Industry 5.0 through the conceptual framework of the Intelligent Cyber-Social Ecosystem—an integrated, human-centered model that emphasizes the convergence of technological innovation, social interaction, and environmental sustainability. In contrast to the automation-driven focus of Industry 4.0, this study offers a systemic perspective in which technologies such as Artificial Intelligence, the Internet of Things, cyber-physical systems, and Big Data are employed not merely for efficiency, but in the service of human creativity, ethical governance, and socio-environmental responsibility. Through a content-based synthesis of relevant literature, the study develops an architectural model of Industry 5.0, presenting an overview of its technological elements, design principles, and intelligent components. These technologies enable personalized manufacturing, intelligent supply chains, circular economy models, and enhanced collaboration between humans and machines. The findings underscore the transformative potential of Industry 5.0 to reimagine industrial development by aligning digital transformation with human values and sustainable practices. The study contributes to the theoretical foundation of Industry 5.0 and offers practical guidance for strategic decision-making in industrial policy, innovation management, and organizational design. It also acknowledges the limitations associated with the early stage of development in this field and outlines directions for future research, including empirical validation, ethical assessment frameworks, and governance models based on stakeholder participation.
Keywords – Industry 5.0, Digital society, Human centric, Artificial intelligence, Social responsibility.

Distributed Learning for Next-Gen Recommender Systems: A survey

Rares Tamaian[1], Voicu Babiciu[1], Oliviu Matei[1], Sanaz Nikghadam[2]
[1]Technical University of Cluj-Napoca, Cluj-Napoca, Romania
[2]NOVA University Lisbon, CTS-UNINOVA and LASI, Portugal
rares.tamaian@campus.utcluj.ro, voicu.babiciu@campus.utcluj.ro, oliviu.matei@gmail.com, sanaznik@uninova.pt

Abstract – The exponential growth of data and model complexity has rendered traditional centralized recommender system approaches insufficient, necessitating the adoption of distributed learning paradigms. This article provides a comprehensive survey of distributed learning algorithms, focusing on their fundamental principles, architectural variations, and critical applications within recommender systems. We delineate the distinctions between centralized, distributed, and decentralized systems, highlighting how distributed approaches address challenges such as computational bottlenecks, inherently distributed datasets, and evolving privacy concerns. The survey explores key distributed learning paradigms, including parallel, federated, and decentralized learning, demonstrating their progressive evolution from optimizing computational speed to ensuring data privacy, robustness, and sovereignty. Specifically, we delve into the application of these algorithms in recommender systems, examining techniques like distributed matrix factorization for scalability, federated learning for privacy-preserving personalization, and parallel distributed training for complex hybrid models. The report details how these methods enable recommender systems to handle massive datasets, provide real-time updates, and safeguard sensitive user information. Finally, we discuss the ongoing challenges and promising future directions in this dynamic field, emphasizing the continuous pursuit of enhanced communication efficiency, robustness to data heterogeneity, advanced privacy mechanisms, and decentralized trust for building more resilient and ethically sound systems.
Keywords – Distributed Learning, Recommender Systems, Federated Learning, Decentralized Learning, Parallel Learning.

Evaluating Large Language Models for Diacritic Restoration in Romanian Texts: A Comparative Study

Mihai Nadăș, Laura Dioșan
Babeș-Bolyai University, Cluj-Napoca, Romania
mihai.nadas@ubbcluj.ro, laura.diosan@ubbcluj.ro

Abstract – Automatic diacritic restoration is crucial for text processing in languages with rich diacritical marks, such as Romanian. This study evaluates the performance of various large language models (LLMs) in restoring diacritics in Romanian texts. Utilizing a comprehensive corpus, we tested models including OpenAI's GPT-3.5, GPT-4, and GPT-4o, Google's Gemini 1.0 Pro, Meta's Llama 2 and 3, MistralAI's Mixtral 8x7B Instruct, Deepinfra's airoboros 70B, and OpenLLM-Ro's RoLlama 2 7B, across different prompt templates ranging from zero-shot to complex multi-shot instructions. Our findings indicate that models such as OpenAI's GPT-4o achieve high diacritic restoration accuracy, significantly surpassing a baseline echo model. However, other models, specifically those from Meta's Llama family, showed varied performance, highlighting the impact of model architecture and training data on task-specific outcomes. This research underscores the need for specialized finetuning and model enhancements to improve NLP tasks involving diacritic-rich languages, providing valuable insights for future developments in computational linguistics.
Keywords – LLM, Diacritic Restoration, NLP, text processing.

Comparative Analysis of Edge Computing Architectures for IoT Systems: Towards Scalable, Secure and Sustainable Deployments

Daniela Timisică[1,2], Radu Boncea[1], Sebastian Balmus[1], Bogdan Dura[1]
[1]National Institute for Research and Development in Informatics - ICI Bucharest
[2]National University of Science and Technology POLITEHNICA Bucharest
daniela.timisica@ici.ro, radu.boncea@ici.ro, sebastian.balmus@ici.ro, Bogdan.dura@ici.ro

Abstract – The proliferation of Internet of Things (IoT) devices has accelerated the demand for efficient, scalable, and secure data processing solutions. While cloud computing traditionally supported large-scale analytics, its limitations in latency, bandwidth, and privacy prompted the rise of edge computing as a decentralized alternative. This paper provides a comparative survey of edge computing architectures tailored for IoT environments, uniquely introducing a multi-layered taxonomy that encompasses far-edge, mid-edge, and near-edge sub-layers, alongside paradigms such as fog computing, cloudlet, and mobile edge computing (MEC). Unlike prior surveys that focused primarily on latency or deployment topologies, this study conducts a holistic analysis that simultaneously considers scalability, security, and sustainability trade-offs, supported by real-world use cases across healthcare, agriculture, manufacturing, and transportation. Intended for IoT system designers, network architects, and researchers in distributed computing, this survey highlights key architectural trade-offs and outlines future research directions for building optimized, intelligent edge-IoT systems.
Keywords – Edge Computing, Internet of Things (IoT), Fog Computing, Cloudlet, Mobile Edge Computing (MEC), Edge architecture, Edge AI, Lightweight virtualization, Low latency, Scalability, Security, Sustainable computing, 5G, Microservices.

Comparative Evaluation of Flow-Based Feature Extraction Tools for IoT Cyberattack Detection

Branly Alberto Martinez Gonzalez, Malena Pérez Sevilla, J.A Rincon Jaime Andrés Rincón Arango, Daniel Urda Muñoz
Grupo de Inteligencia Computacional Aplicada (GICAP), Departamento de Digitalizaci'on, Universidad de Burgos, Burgos, Spain
{bamgonzalez,mpsevilla,jarincon,durda}@ubu.es

Abstract – The increased deployment of Internet of Things (IoT) devices has led to greater complexity and scale in network traffic, introducing challenges in effectively detecting cyber threats. This study compares two network flow feature extraction tools, NFStream and Tranalyzer, through their performance in detecting malicious network traffic using a Multilayer Perceptron (MLP) neural network. The evaluation was conducted using datasets generated from laboratory-based IoT setups involving edge devices, such as the Raspberry Pi 4 and Jetson Nano, as well as data from the publicly available ToN-IoT dataset. Each tool independently processes the same raw network captures, creating two distinct feature sets. Both datasets underwent identical preprocessing steps, including balanced undersampling, categorical encoding, and feature scaling, to ensure fair comparison. MLP neural networks, implemented using the Keras Functional API, were trained, validated, and tested on each dataset separately. The results, represented by global and network-specific confusion matrices, demonstrate the high accuracy of both tools, with subtle but measurable differences in performance. Both NFStream and Tranalyzer demonstrated strong and consistent performance under identical evaluation settings, confirming their capability to support machine learning-based detection of cyberattacks in IoT environments.
Keywords – IoT, Cybersecurity, Network Traffic Analysis, Feature Extraction, Multilayer Perceptron (MLP), Network Anomaly Detection.

Multi-Quantum Agent System for DNA Self-Assembly

Vasile Vancea
University of Bucharest, Bucharest, Romania
University of Bucharest, Bucharest

Abstract – DNA nanotechnology provides a versatile foundation for designing programmable self-assembling structures, but optimal control over such molecular systems remains a challenge. This paper introduces a novel Multi-Quantum Agent System (MQAS) to coordinate and enhance DNA self-assembly processes. The proposed model integrates quantum-enhanced agents, classical sensing, reinforcement learning, and DNA nanostructure environments. By leveraging quantum circuits for decision-making and entangled communication for coordination, MQAS agents can dynamically interact with the DNA assembly grid to guide structure formation. We detail the system architecture, agent behavior, quantum learning mechanisms, and simulation environment, and compare our framework to classical multi-agent DNA systems. Experimental scenarios and future directions demonstrate the potential of MQAS to advance programmable matter and synthetic biology.
Keywords – Multi-Quantum Agent System, DNA Self-Assembly.

Quantum Clustering for Classification of DNA Sequences

Vasile Vancea
University of Bucharest, Bucharest, Romania
University of Bucharest, Bucharest

Abstract – Recent advances in quantum machine learning (QML) have demonstrated the potential of quantum algorithms to tackle complex classification tasks beyond the capabilities of classical methods. In this study, we propose a novel quantum clustering (QC)-based framework for the classification of DNA sequences. Building upon methods of QC, we extend the paradigm to biological sequence data by encoding nucleotide sequences into a quantum-compatible feature space using one-hot, k-mer, and amplitude encodings. Our approach combines quantum evolution under a potential landscape de-fined by data density with a hybrid quantum-classical variational classifier to distinguish between DNA sequence classes. We implement the full pipeline on near-term quantum hardware using variational quantum circuits and simulate performance using Qiskit and PennyLane frameworks. Our method demonstrates robust performance in classifying synthetic and biological DNA datasets, achieving competitive accuracy with enhanced clustering resolution in high-dimensional quantum Hilbert spaces. This work represents the application of QC to DNA classification and highlights the promise of quantum-enhanced bioinformatics in the NISQ era.
Keywords – Quantum Machine Learning, DNA Sequences Classifications, Quantum Clustering.

Applications

Free-range chicken farms oriented bird detection

Enol García González[1], José Ramón Villar[1], Javier Sedano[2]
[1]University of Oviedo, Computer Science Department, Oviedo, Spain
[2]Instituto Tecnológico de Castilla y León, Burgos, Spain
{garciaenol,villarjose}@uniovi.es, javier.sedano@itcl.es

Abstract – This work detects predators in extensive breeding farms, especially those dedicated to free-range chicken production. In this environment, the main threats are birds of prey and poachers. The proposal focuses on the first case to study how using deep learning models can help detect and label images. Specifically, we evaluate the performance of several object detection models on a dataset focused on birds of prey in outdoor farm settings. The study compares architectures from two families: Faster R-CNN and YOLO. Results show notable differences in performance between model families. Faster R-CNN shows a better performance than YOLO. This considerable difference marks the importance of selecting the model to be used, especially in cases such as this one, where the objects to be detected and classified are small in size.
Keywords – Object detection, Bird detection, Faster R-CNN, YOLO, Wildlife monitoring.

Innovative Medical Image Analyzer - iMIA tested on COVID-19 images

Cristina Ticala[1], Camelia-M. Pintea[1], Oliviu Matei[1], Ricardo Peres[2]
[1]Technical University of Cluj-Napoca, Cluj-Napoca, Romania
[2]NOVA University Lisbon, CTS-UNINOVA and LASI, Portugal
cristina.ticala@mi.utcluj.ro

Abstract – This paper presents an innovative and integrative application designed for the preprocessing and analysis of medical images, with a particular focus on edge detection and image filtering. The application aims to support diagnostic processes by enhancing image features critical to clinical interpretation. The actual version of the *innovative Medical Image Analyzer* (*iMIA*) platform includes two primary functional modules: Edge algorithms, which implements both classical and artificial intelligence related edge detection techniques, and Filters & Analysis, which applies spatial filters to improve contrast and reduce noise. By combining these techniques, the tool facilitates the identification of anatomical boundaries and regions of interest in medical scans, particularly CT images. COVID-19 Lung CT medical images were used to test the application and delivered promising results. The application provides an intuitive interface and modular structure, making it suitable for both educational and research purposes in medical image processing.
Keywords – Medical image analysis, COVID-19, Image edge detection.

DeepFlow – Deep Clustering for Blood Cell Classification and Quantification

Mihaela Macarie-Ancau, Adrian Groza
Department of Computer Science, Technical University of Cluj-Napoca, Romania
Macarie.Io.Mihaela@cs.utcluj.ro, Adrian.Groza@cs.utcluj.ro

Abstract – Accurate classification of blood cells plays a key role in improving automated blood analysis for both medical and veterinary applications. This work presents a two-stage deep

clustering method for classifying blood cells from high-dimensional signal data. In the first stage, red blood cells (RBCs) and platelets (PLTs) are separated using a combination of an improved autoencoder and the IDEC algorithm. The second stage further classifies RBC subtypes, pure RBCs, reticulocytes, and clumped RBCs, through a variational deep embedding (VaDE) approach. Due to the lack of detailed cell-level labels, soft classification probabilities are generated from sample-level data to approximate the true distributions. The aim is to contribute to the development of low-cost, automated blood analysis systems suitable for veterinary and biomedical use. Initial results indicate this method shows promise in effectively distinguishing different blood cell populations, even with limited supervision.
Keywords – Deep Clustering, Flow Cytometry, Blood count Instruments, Veterinary Hematology.

Exploratory Analysis and Classification of Lymphoma DNA Methylation Data

Maria Iuliana Bocicor[1], Thomas Erblich[2]
[1]Department of Computer Science, Faculty of Mathematics and Computer Science, Babes-Bolyai University, Cluj-Napoca, Romania
[2]Queen Elizabeth Hospital, Lewisham and Greenwhich NHS Trust, Stadium Road, London
maria.bocicor@ubbcluj.ro, thomas.erblich@nhs.net

Abstract – Aberrations in DNA methylation patterns are associated with various diseases, including cancer, and their study may provide valuable signs for early detection of this disease. The aim of this paper is to propose a methodology for the analysis of such epigenetic profiles, with the objective of distinguishing between different lymphoma subtypes. In our analysis, we start from DNA methylation data, we identify differentially methylated regions using a proposed naive approach and an established statistical approach (Dispersion Shrinkage for Sequencing algorithm), perform feature selection using variance thresholding and the Minimum Redundancy Maximum Relevance method and apply non-linear feature extraction with autoencoders. Lastly, we train three classifiers (XGBoost, Support Vector Machines, and Artificial Neural Networks) on the obtained, reduced feature sets. The results show that a particular combination of DMR extraction procedure, feature selection/extraction and classifier performs best on both an internal test set (held-out from the original data) and an external test set. However, considering the limited size of the input data (only 30 samples), these findings should be considered preliminary and require further validation in future studies.
Keywords – Methylation data, Lymphoma subtypes, Machine learning.

Applying L-Moments Statistical Theory to Analyze Real-World Sensor Data from Orchards

José Ramón Torres-Martín, Yolanda Carrión-García, José Manuel Velarde-Gestera, Inmaculada Mora-Jiménez, Mihaela I. Chidean
[1]Escuela Internacional de Doctorado, Universidad Rey Juan Carlos, Madrid, Spain
[2]Drimay Consultores S.L., Seville, Spain
[3]Deptartament of Signal Theory and Communications, Telematics and Computing Systems, Universidad Rey Juan Carlos, Madrid, Spain
[4]Data, Complex Networks and Cybersecurity Research Institute, Universidad Rey Juan Carlos, Madrid, Spain
jr.torres.2023@alumnos.urjc.es, {josetm,jmvelarde}@drimay.es,
{yolanda.carrion, inmaculada.mora, mihaela.chidean}@urjc.es

Abstract – The increasing demand for food and climate change are challenges for the agricultural sector. The integration of IoT systems and the use of machine learning enable smart irrigation system and optimize the use of water. In this study, we use the L-moments statistical

theory in an exploratory analysis aiming to extract distribution characteristics of variables related to evapotranspiration, specifically the ambient temperature, the atmospheric pressure, the soil moisture and the wind speed. In our analysis we use actual data collected between 2022 to 2025 by IoT sensors installed in 20 orchards located in different Spanish regions. Our results reveal the possibility for each considered variable to distinguish regions with different climatical conditions, which is crucial decision-making for irrigation plans and optimization problems.

Keywords – Agricultural Sector, Internet of Things (IoT), Smart Irrigation, L-moments, Moment Ratio Diagram.

L-Moments for Robust Temporal Data Drift Monitoring in Agricultural IoT Systems

Yolanda Carrión-García, José Ramón Torres-Martín, Inmaculada Mora-Jiménez, José Manuel Velarde-Gestera, Mihaela I. Chidean

[1]Escuela Internacional de Doctorado, Universidad Rey Juan Carlos, Madrid, Spain

[2]Drimay Consultores S.L., Seville, Spain

[3]Deptartament of Signal Theory and Communications, Telematics and Computing Systems, Universidad Rey Juan Carlos, Madrid, Spain

[4]Data, Complex Networks and Cybersecurity Research Institute, Universidad Rey Juan Carlos, Madrid, Spain

{yolanda.carrion, inmaculada.mora, mihaela.chidean}@urjc.es, jr.torres.2023@alumnos.urjc.es, {josetm,jmvelarde}@drimay.es

Abstract – Data-driven models are increasingly used in agriculture to optimize irrigation and promote efficient water use. However, their reliability depends on the stability of the underlying data distributions and data drift, which can be affected by data drift induced by seasonal variability, evolving climatic conditions, or extreme events. In this work, we propose a methodology for constructing confidence intervals around statistical moment estimates to characterize the temporal evolution of variable distributions and identify significant shifts in their underlying dynamics. The approach leverages L-moment statistical theory, with particular emphasis on higher-order L-moments (L-skewness and L-kurtosis), to detect data drift in key environmental variables. We assess the performance of L-moments against classical product moments using real-world data collected between 2022 and 2025 by IoT sensors deployed in 20 orchards across various Spanish regions. Results show that L-moments yield more stable and robust estimates under the varying conditions of these variables and also for small sample sizes, and exhibit greater resilience to outliers. This enables more reliable drift detection without mistaking noise or anomalies for actual distributional changes, which is essential for maintaining robust and adaptive models in agricultural systems. While the application of the proposed methodology is illustrated through a smart farming application, it is readily generalizable to other domains.

Keywords – Data stability, Moment-based drift detection, L-moments, Classical moments, Smart farming, Sensor networks.

Forecasting technical debt in software projects with limited historical data

Oskar Picus

Department of Computer Science, Faculty of Mathematics and Computer Science, Babeș-Bolyai University, Cluj-Napoca, Romania

oskar.picus@ubbcluj.ro

Abstract – Technical debt refers to the compromises made during software development that enable rapid delivery but may hurt the long-term health of a software system. Recently,

forecasting technical debt has become a topic of interest to researchers and techniques for it were used successfully for mature software projects. This work aims at analyzing if existing techniques in forecasting technical debt are still applicable and effective for software projects with limited historical data. We have partially applied the methodology of a reference study, using the sliding-window method and various Machine Learning algorithms, including a boosting one, which was not used before for forecasting technical debt. These techniques were applied on a data set from three projects, 62% to 74% smaller than the one in the reference study. We have obtained comparable results, however, we observed a general inability of existing Machine Learning techniques of coping with small and irregular data sets. This research contributes to the ongoing effort to improve software maintainability and offers insights to practitioners on the methods for mitigating the adverse effects of technical debt. Our results partially replicated the original study's, emphasizing the need for further research aimed at forecasting technical debt for software projects with minimal data available.
Keywords – Technical debt, Time series, Regression, Forecasting.

Technical Debt, empirical study, Human Aspects in Software Engineering

Ciprian-Viorel Stupinean
Babes-Bolyai University Cluj-Napoca, Romania
ciprian.stupinean@ubbcluj.ro

Abstract – Technical debt (TD) is a widespread issue in software development, often due to poor understanding and management. This qualitative study used a survey and thematic analysis to understand how developers identify and manage TD from individual perspective but also as part of development teams. We found that developers use techniques like code reviews, testing, backlog activities, and team communication. Our findings emphasize that managing TD effectively requires both individual effort and team collaboration, suggesting a need for a holistic, cross-project approach considering culture, tools, recurring issues, and collective solutions.
Keywords – Technical Debt, Developer Perspective, Qualitative Study.

Analyzing Similarity Across Investment Portfolios: A Multi-Dimensional Approach

Maria Oikonomidou, Ferad Zyulkyarov
inbestMe, Barcelona, Spain
{maria.oikonomidou,ferad.zyulkyarov}@inbestme.com

Abstract – This paper presents a synthetic dataset and evaluation framework for analyzing investment portfolio similarity. Understanding how portfolios relate in terms of structure and performance behavior – such as risk, return, and sensitivity to market movements – is essential for strategy comparison, benchmarking, and risk-based clustering in both research and industry. We generate 1,000 portfolios using a diverse asset universe, compute a rich set of performance and risk metrics, and represent portfolio relationships through similarity graphs. Four similarity functions – Jaccard, Cosine, Euclidean, and Pearson correlation – capture different dimensions of portfolio composition and performance behavior. Using hierarchical clustering and standard evaluation metrics, we show that behavior-based similarities, particularly cosine similarity, produce the most interpretable clusters. Our contributions include a reproducible methodology for synthetic portfolio generation, a set of similarity

metrics, and a framework for evaluating clustering effectiveness. This resource supports further research in portfolio modeling, recommendation systems, and financial strategy design.
Keywords – Portfolio similarity, Performance metrics, Risk analysis, Graph representation, Financial benchmarking.

Survey on Cybersecurity in Precision Agriculture: Threat Landscape, Defense Strategies, Ethics and EU Regulations

Valentin-Alexandru Vladuta[1], Emil Marian Pasca[1], Javad Jassbi[2]
[1]Technical University of Cluj-Napoca, Electric, Electronic and Computer Engineering Department, Bucharest, Romania
[2]NOVA University Lisbon, CTS-UNINOVA and LASI, Portugal
vladuta.alexandru.valentin@gmail.com, emil.pasca@campus.utcluj.ro, jassbi@uninova.pt

Abstract – Precision agriculture has emerged as an important topic in different contexts, as it is considered a potential solution to some of the most critical global challenges: global warming and its effects on crops, cyber attacks on food supply chains, decreasing availability in labor, and the list continues. The scope of the current paper is to make a survey on security and ethical considerations of precision agriculture, to understand the phenomenon behind it: from background, to challenges and potential solutions. It is critical to discover what can become an attack surface and generate vulnerabilities, by studying in-depth patterns, trends, standards, and best practices. Only then can one propose solutions to existing challenges and offer guidelines for building a system following the secure-by-design principle. In addition, ethics and regulations in precision agriculture are considered, as specific literature does not cover this perspective, enabling the current study to fill an important gap. In conclusion, technology plays an important role in all relevant fields today, and agriculture is no exception, as the food industry will remain one of the pillars of global security and development. This survey identifies three critical gaps: regulatory-technical implementation challenges, vulnerability cascading traceability issues, and human-centered security vulnerabilities specific to agricultural contexts.
Keywords – Attack vectors, Cybersecurity, Defense strategies, Ethics, Precision agriculture, Regulations, Vulnerabilities.

Association Rule Mining for Combinatorial Biomarker Discovery. A Case Study on Type 2 Diabetes

Ioana Ciuciu[1], Dana Ciobanu[2], Daniel Murg[1], Cosmin Lazar[1]
[1]Babes-Bolyai University, Cluj-Napoca, Romania
[2]Iuliu Hatieganu University of Medicine and Pharmacy, Cluj-Napoca, Romania
dana.ciobanu@umfcluj.ro

Abstract – The article investigates the application of association rule mining, in particular the Frequent Pattern Growth (FP-Growth) algorithm, for combinatorial biomarker discovery for Type 2 Diabetes disease. The dataset used for the case study contains measurements from 100 subjects. The results obtained identify “interesting” rules associated with Type 2 Diabetes which were evaluated by the domain expert as being 88,23% relevant and 11,77% partly relevant. A comparative study aiming to compare rule mining methods with other techniques for biomarkers discovery approaches is work in progress.
Keywords – Association Rule Mining, Combinatorial Biomarker Discovery, FP-Growth, Type 2 Diabetes.

Optimizing Diabetes Risk Prediction: Comparative Analysis of Machine Learning Models and Feature Importance

Ana-Maria Păpară
Faculty of Mathematics and Computer Science, Babeș-Bolyai University, Cluj-Napoca, Romania
ana.maria.papara@gmail.com

Abstract – Diabetes is known to be one of the most common diseases in the world today. A significant alternative to traditional diagnostic methods that can be time and resource-consuming is an ML-based approach. With the use of historical health lifestyle and diagnostic data, artificial intelligence can predict someone's risk of developing diabetes. This study aims to explore the potential of ML models for the prediction of diabetes based on different health and lifestyle factors. Furthermore, using a feature importance analysis, the objective is to determine the factors that affect the most diabetes condition. By evaluating multiple methods, configurations and datasets, current study aims to identify the most effective approach for predicting diabetes risk. For training and prediction, the following 4 models were utilized: Random Forest Classifier (RFC), LightGBM, Decision Tree (DTC) and Naive Bayes. The performance of the models was evaluated on the basis of accuracy, precision, recall and F1-score. The study involved testing these models on two datasets: a smaller one that contains medical attributes (PIMA dataset) and another much larger that incorporates mostly lifestyle factors (derived from the Behavioral Risk Factor Surveillance System Survey from 2022). The results show that the LightGBM and Random Forest models outperform the other evaluated classifiers. LightGBM achieved the highest accuracy of 86% in the larger dataset, highlighting the effectiveness of incorporating lifestyle factors. Furthermore, feature importance analysis revealed key predictors, such as weight, height and mobility difficulties to be those that impact diabetes diagnostic the most.
Keywords – Diabetes, ML, Feature importance, Prediction.

Assessment of groundwater quality of the Republic of Ireland using Random Forest algorithm

Tahmida Chowdhury, Rajat Nag, Md Arman Habib , Md Salauddin
University College Dublin, Dublin, Ireland
tahmida.n.chowdhury@ucdconnect.ie, raj.nag@ucd.ie, md.habib@ucdconnect.ie, md.salauddin@ucd.ie

Abstract – Water is considered the most important element for humans and other living organisms. However, escalating pollution and the impacts of climate change necessitate the analysis of groundwater quality, which is imperative for ensuring safe water. Due to its scalability, robustness, and reduced susceptibility to overfitting, the Random Forest (RF) classifier method was used to calculate the water quality index for the classification of water quality. The groundwater quality in the research area is forecasted by constructing a Random Forest model utilising the arithmetic Water Quality Indices (WQI), which encompass the key elements affecting groundwater quality, including pH, temperature, nitrate, phosphorus, and other chemical parameters. The F1 score, accuracy, recall, and precision are used to evaluate the performance. The model achieved a satisfactory performance, with an overall accuracy of 96%, an F1 score of 99% for the excellent class, and a score of 97% for the unsuitable class. In addition, excellent water quality constitutes 55% of the samples; however, concerns persist for poor and unsuitable classes, which comprise 10% and 16% of the samples, respectively. Furthermore, the feature analysis revealed that the most important variables influencing the quality of the groundwater in the study region are Escherichia coli (E. coli) and phosphorus. The study provided significant insights into the groundwater quality of the study area,

establishing a basis for future research and the development of effective groundwater management policies.

Keywords – Groundwater, Random Forest, Water Quality Index, Pollution, Water Resources Management.

AI-based Sensor Optimization and Anomaly Detection in Water Distribution Networks

Furqan Rustam[1], Usman Saeed[1], Anca Delia Jurcut[1], Gabriele Freni[2], Mariacrocetta Sambito[2], Md Salauddin[1]

[1]School of Computer Science, University College Dublin, Ireland

[2]School of Engineering and Architecture, University of Enna "Kore", Cittadella Universitaria, Italy

furqan.rustam@ucdconnect.ie, mohd.u.khan@ucdconnect.ie, gabriele.freni@unikore.it, mariacrocetta.sambito@unikore.it

Abstract – *Water Distribution Systems* (WDSs) are critical infrastructures that rely on advanced technologies, such as IoT devices and smart sensors, to ensure efficient water management and delivery. While these innovations enhance operational reliability, they also increase the complexity of monitoring by adding more devices, making the network more resource-intensive and susceptible to cyberattacks. Strategic sensor placement and optimization are essential to maintaining robust anomaly detection while minimizing the number of monitored devices. In this study, we propose a sensor placement optimization and anomaly detection framework for WDSs, aimed at evaluating the effectiveness of water pumps in their current positions for anomaly detection. Using the BATADAL dataset, which contains data from water pumps, junctions, and tanks, we analyse the contribution of each pump's location to anomaly detection. Our approach employs the Gini index within a Random Forest model to rank the importance of water pumps and assess whether their placement contributes effectively to the monitoring system. Pumps that provide minimal utility for anomaly detection are flagged for exclusion, allowing for the optimization of resource allocation. Further, we integrate a *Long Short-Term Memory* (LSTM) model to predict anomalies using only the essential devices. The LSTM model employs a continuous learning mechanism, enabling it to adapt to dynamic changes in the WDS, such as the addition or removal of sensors. This adaptability ensures the system remains reliable in real-time operations, maintaining robustness even in evolving scenarios.

Keywords – Continuous Learning, Resource Optimization\and Sensor Placement\and Water Distribution Networks.

Multi-View Sky Image Regression for Enhanced Photovoltaic Power Forecasting

Alexandru-Adrian Ciobanu[1], Fereshteh Jafari[2,3], Joseph Moerschell[2], Ariana-Andra Șerpar[1,4,6], Nebojsa Bacanin[5], Catalin Stoean[1]

[1]Department of Computer Science, University of Craiova, Craiova, Romania

[2]Institute of Systems Engineering, University of Applied Sciences and Arts Western Switzerland, Sion, Switzerland

[3]Institute of Computer Science, University of Bern, Switzerland

[4]Department of Computer Science, West University of Timișoara, Timișoara, Romania

[5]Faculty of Informatics and Computing, Singidunum University, Belgrade, Serbia

[6]Department of Computer Engineering, Automation and Robotics, University of Granada, Granada, Spain

ciobanu.alexandru.adrian99@gmail.com, catalin.stoean@inf.ucv.ro, fereshteh.jafari@hevs.ch, joseph.moerschell@hevs.ch, fereshteh.jafari@unibe.ch, ariana.serpar00@e-uvt.ro, nbacanin@singidunum.ac.rs, arianaserpar34@correo.ugr.es

Abstract – Accurate photovoltaic (PV) production forecasting is essential for optimizing energy management in modern power systems. While existing methods often rely on time

series analysis or meteorological data, this study explores the possibility of treating PV forecasting as a regression problem using sky images, without benefiting from information within temporal dependencies. We evaluate whether contrast and brightness features extracted from sky images captured simultaneously, but with different exposure settings, can provide sufficient predictive power for PV production. Three key scenarios are investigated: (1) regression on individual image types, (2) the impact of including the hour-of-day as a contextual feature, and (3) the fusion of contrast and brightness features across multiple image types. Using 5-fold cross-validation and a diverse set of 9 regression models (linear, tree-based, and ensemble methods), we demonstrate that multi-modal data integration, i.e., combining features from multiple images with the hour-of-day, significantly improves predictive accuracy.
Keywords – Photovoltaic forecasting, regression analysis, sky images, feature fusion, ensemble models.

Forecasting Higher Education Trends Using Multi-Output Random Forest Regression and Regional Economic Weighting: A Case Study Based on Romania

Bogdan Vaduva[1], Laura Andreica[1], Mara Diana Hajdu-Macelaru[2], Anca Avram[1], Erina Guraziu[3]

[1]Department of Electrical Engineering, Electronics and Computers, Technical University of Cluj-Napoca, Str. Victor Babes nr. 62/A, Baia Mare, Maramureș, Romania

[2]Department of Mathematics and Informatics, Technical University of Cluj-Napoca, Str. Victoriei, nr. 76, 430122, Baia Mare, Maramures, Romania

[3]OpenCom issc, Piazza Giotto, 13 - 52100 Arezzo (AR), Italy

bogdan.vaduva@ieec.utcluj.ro, mara.macelaru@cunbm.utcluj.ro, erina.guraziu@opencom-italy.org

Abstract – This work investigates a novel approach to forecasting higher education trends by integrating a multi-output Random Forest regression model with a weighting system that captures the influence of different university centers. In addition to forecasting key indicators, such as faculty, total student enrollment and teaching staff, the methodology creates an influence map that illustrates the relative impact of these centers on the overall educational landscape. The model uses a Multi Output Regressor to predict multiple interrelated targets while incorporating influence weights. Performance is evaluated using metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2). Forecasts extending 10 years into the future indicate high model accuracy, with R^2 values near unity for all variables. The influence map created from the weighted predictions provides insights for strategic planning and resource allocation in higher education. This research contributes to the growing body of knowledge on predictive analytics in education by offering a framework that combines advanced machine learning with visualization of institutional impact.
Keywords – Specialization, Machine Learning, Forecast, Random Forest, Multi-output regression, ARIMA, Time series prediction, Educational planning, Influence map, Regional weighting, Predictive analytics, Higher education policy.