SACAIR2024

Southern African Conference for
Artificial Intelligence Research

AI for Societal Impact

University of the Free State, Bloemfontein

2 - 6 December 2024

Book Of Abstracts

Contributions to the Unconference of the 5th Southern African Conference for Artificial Intelligence Research (SACAIR) 2024

Editors:
Lurgasho Minnie
Jacques Maritz
Armand Bester
Stefaans Erasmus
Albert Van Eck

SACAIR
Southern African Conference for Artificial Intelligence Research

Southern African Conference for
Artificial Intelligence Research

SACAIR 2024

Preface

The unit of Engineering Sciences at the University of the Free State are extremely honoured to host the fifth Southern African Unconference for Artificial Intelligence Research (SACAIR) 2024. SACAIR is associated with the Centre for Artificial Intelligence Research (CAIR) which was founded in 2011 and is a research network in South Africa that aims to develop high-quality Artificial Intelligence research expertise in the country. The use of Artificial Intelligence (AI) has indeed taken the world by storm. AI has shown prominence in the research industry and has accelerated discoveries in a wide range of research fields. The theme of this year's SACAIR conference will be on the societal impact of artificial intelligence. This theme was chosen because the progress made from artificial intelligence are usually not realized in our current situations, especially in regions facing ongoing societal struggles involving various factors like politics, economics, history, and the environment. This conference aims to bring together national and international researchers from a wide range of disciplines including Physics, Statistics, Computer Science, Mathematics, Informatics, Philosophy and Humanities. This will foster connections between established and upcoming researchers that uses artificial intelligence in their research. A call for abstracts was announced and 15 abstracts from students specializing in various fields were received. From these abstracts, 13 will be in a presentation format and the remaining 2 will be in a digital poster format that will take place on the 2nd of December. A day for tutorials will take place on the 3rd of December 2024 followed by the main conference on the 4-6 of December 2024. The student unconference committee would like to dearly thank all the researchers that submitted abstracts for SACAIR 2024. From these abstracts, it is evident that artificial intelligence will be used to better humanity and propel us forward in the everlasting age of discovery. Finally, we would like to give a special thanks to our partners and sponsors of SACAIR 2024. Without their support this event would not be possible!

Thankful for our sponsors!

UNIVERSITY OF THE
FREE STATE
UNIVERSITEIT VAN DIE
VRYSTAAT
YUNIVESITHI YA
FREISTATA
UFS

Artificial
Intelligence
www.elsevier.com/locate/artint

cair
CENTRE FOR ARTIFICIAL
INTELLIGENCE RESEARCH

NITheP
National Institute for
Theoretical Physics

NITheCS
National Institute for
Theoretical and
Computational Sciences

Table of Contents

Critiquing Cross-Cultural Ethics in Artificial Intelligence in Education (AIED)

Author: Nasreen Watson, University of Johannesburg, Philosophy Department

Abstract:

Ubuntu, an African ethical principle advocating for a communal way of life and interconnectedness among citizens in South Africa, has been purported to apply to all areas of social and economic spheres. However, challenges persist in the development and deployment of Artificial Intelligence in Education (AIED), which includes a consensual agreement on ethical frameworks to guide practical usage in supporting first-year university students. In response to these challenges, Enslin and Horsthemke (2004) argue that Ubuntu's tenets should be considered universal because of the overarching similarity with other humanistic philosophical approaches. They argue that the prioritization of Ubuntu principles in education fails to recognise common value structures and principles from Western ethical influence, thereby challenging Ubuntu's effectiveness as the primary ethical framework in AIED. This paper addresses the central question: What tensions persist in the practical application of Artificial Intelligence in Education (AIED) within a cross-cultural ethical framework? My aim is to expand upon the arguments of Enslin and Horsthemke (2004) by critiquing the constituents of 'citizenship education' and its ethical framework in Ubuntu as the underpinning principle of African democracy. Although recent efforts by African philosophers have attempted to establish various ethical frameworks for AI, the application and implementation of these frameworks within the educational sector remain unclear. Moreover, current literature commits the 'fallacy of equivocation'; ethical concepts such as 'Ubuntu' are used to assert that successful ethical frameworks of AIED can only be achieved through the critique of Eurocentrism. However, this focus hinders the advancement of African educational institutions to adopt and implement a global standard of partnership that could advance student development.

Keywords: Ubuntu, Artificial Intelligence in Education (AIED), Citizenship Education, Western Philosophy, Eurocentric

Exploring Predictive Anomaly Detection in Mineral Processing: A Machine Learning Approach

Morne C. Du Plessis[1], Deshendran Moodley[1,2]

[1] University of Cape Town, Cape Town, South Africa
[2] Centre for Artificial Intelligence Research (CAIR), South Africa
dplmor006@myuct.ac.za, deshen@cs.uct.ac.za

Abstract. Operational efficiency and safety are paramount in the minerals processing industry, where detecting abnormal conditions before they result in costly downtimes or equipment failures will be highly beneficial. This paper explores the application of advanced machine learning (ML) techniques for predictive anomaly detection in minerals processing plants. Using a real-world dataset from a South African facility, which includes sensor data such as vibration and power draw, we apply an automated labelling. To predict failures, we propose employing a neural network trained on the labelled data to recognize patterns that indicate potential system failures within a one-step prediction horizon. By effectively using predictive ML-based anomaly detection methods, we demonstrate how ML facilitates the transition from reactive to proactive maintenance strategies.

Keywords: Predictive Anomaly Detection, Machine Learning, Neural Networks, Minerals Processing

1 Introduction

The study focuses on developing and evaluating various machine learning models that can proactively detect anomalies in autogenous grinding (AG) mill operations. These models aim to forecast deviations from normal operating conditions (NOC), thereby facilitating the early identification of abnormal operating conditions (AOC) and preventing system failures as defined by V. Groenewald et al. (2006). This paper presents an exploration of statistical based automatic labeling and machine learning techniques for predictive anomaly detection in mineral processing plants.

One of the key contributions of this research is the integration of forecasting into the anomaly detection system, allowing the models to predict anomalies before they occur. This proactive approach contrasts with traditional anomaly detection systems that merely identify anomalies when or after they have manifested. Another significant contribution is the real-world application of these techniques in a minerals processing plant.

2 Experimental Design

A real-world dataset from a South African mineral processing facility was used to determine the operating states as well as train and test machine learning models. The dataset spanned over 6-months, containing 5-minute interval readings of vibration and mill power draw.

2.1 Statistical Based Automatic Labeling

We expand the method used by Kulanuwat et al. (2021) where we use percentiles to determine out-of-bound limits for individual sensors inputs. The out-of-bound sensor inputs were aggregated and normalized to distinguish NOC from AOC, forming the foundation for anomaly detection.

2.2 ANN Based Predictive Modeling

The cleaned and labeled data was then preprocess further and the prepared dataset was used to train an Artificial Neural Network (ANN) to predict failures with a 1-step horizon. Walk-forward validation, as suggested by Kouassi and Moodley (2020), was used to ensure continuous model adaptation to new data, simulating real-time deployment.

3 Preliminary Results

The for the real-world dataset, the automated labelling method classified of 98.59% for NOC and 1.08% for AOC. The automated labelling method effectively labeled the anomalous periods leading up to system breakdowns, although it struggled with high variability, sometimes misclassifying short-duration events. Despite this, it proved valuable in identifying and labeling significant disruptions, laying a solid foundation for machine learning predictions.

The ANN achieved an accuracy of 98.23% on the hold-out test set, with a precision of 77.03% and a recall of 52.2%, highlighting its strong ability to detect anomalies. However, the model's focus on high recall led to a higher number of false positives, reflected in a lower precision. Despite this, the ANN successfully predicted abnormal conditions with a 1-step horizon, accurately identifying the lead-up to system failures.

This research offers practical insights into the deployment of ML-based predictive maintenance systems in the minerals processing industry. By shifting from reactive to proactive maintenance, these systems can help optimize operational efficiency, reduce unplanned downtimes, and improve overall safety.

Future work will explore the incorporation of more advanced deep learning architectures, such as Long Short-Term Memory (LSTM) networks, to further refine the system's accuracy and reduce the occurrence of false positives.

References

1. Kulanuwat, L., Chantrapornchai, C., Maleewong, M., Wongchaisuwat, P., Wimala, S., Sarinnapakorn, K., Boonya-aroonnet, S.: Anomaly detection using a sliding window technique and data imputation with machine learning for hydrological time series. Water 13(13), 1862 (2021)
2. de V. Groenewald, J., Coetzer, L., Aldrich, C.: Statistical monitoring of a grinding circuit: An industrial case study. Minerals Engineering 19, 1138–1148 (2006). Featuring selected papers presented at Process Systems for the Metallurgical Industries ’05 symposium, Cape Town, South Africa. https://doi.org/10.1016/j.mineng.2006.05.009
3. K. H. Kouassi, D. Moodley, An analysis of deep neural networks for predicting trends in time series data, in: Proceedings of the South African Conference for Artificial Intelligence Research (SACAIR), Springer, 2020, pp. 119–140. URL: https://doi.org/10.1007/978-3-030-66151-9_8. doi:10.1007/978- 3- 030- 66151- 9_8.

Spatial-Temporal Graph Neural Networks for temperature prediction

Lekuba Ntoane[1] and Deshendran Moodley[1,2]

[1] University of Cape Town, Cape Town, South Africa
[2] Centre for Artificial Intelligence Research (CAIR), South Africa
ntnlek001@myuct.ac.za and deshen@cs.uct.ac.za

Abstract. This research explores Spatial-Temporal Graph Neural Networks (STGNNs) for predicting temperature at South African weather stations. We compare Graph WaveNet (GWN) and Temporal Convolutional Neural Network (TCN), highlighting that GWN captures critical spatial-temporal dependencies over a 48-hour prediction horizon. Network centrality metrics are used to analyse these dependencies between weather stations.

Keywords: Graph Neural Networks · Spatial-Temporal dependencies · Weather prediction · Network centrality metrics.

1 Introduction

Deep Neural Networks (DNNs), including Long Short-Term Memory (LSTM) networks [1][6] and Temporal Convolutional Networks (TCNs) [3], have been effective in weather prediction by focusing on temporal modeling of elements such as temperature, wind speed, and humidity. However, these methods are limited for capturing spatial dependencies between weather stations. Spatial-Temporal Graph Neural Networks (STGNNs) address this limitation by modeling both temporal and spatial dependencies [5]. Davidson et al. [2] demonstrated that STGNN architectures like Weighted Graph Convolution LSTM (WGC-LSTM) [7] and Graph WaveNet (GWN) [8] improve short-term temperature prediction. This research investigates STGNNs for long-term (48-hour) temperature prediction and uses network analysis to analyse the dependencies between weather stations.

2 Experimental Design

The real-world weather dataset used in our experiments was provided by the South African Weather Service (SAWS) taken from 27 weather stations located in different regions of South Africa. Measurements of temperature, rainfall, wind speed, wind direction, humidity, and pressure were obtained from each weather station. Our goal is to predict temperature over a 48 hour prediction horizon. We evaluated and compared the performance of GWN and TCN models over the 48 hour horizon. Finally, we applied the weighted in-degree centrality measure [4] to identify weather stations most influenced by other weather stations.

3 Preliminary Results

We present the preliminary results for temperature prediction over the 48 hour prediction horizon. GWN outperforms TCN, achieving an average MSE of 0.0059 and SMAPE of 12.69%, compared to TCN's average MSE of 0.0139 and SMAPE of 23.65%. Fig 1 shows the performance comparison of TCN and GWN across all weather stations using the SMAPE metric. The best performance from both TCN and GWN is observed at weather station 11 while the worst performance is observed at weather station 20 for TCN and weather station 21 for GWN. Furthermore, GWN constantly outperforms TCN on all weather stations. This improvement in GWN performance over TCN is due to the inclusion of spatial information from neighbouring weather stations.

The spatial-temporal dependencies of GWN are visualised overlaid over the South Africa map in Fig. 2. The weighted in-degree centrality measure is employed to identify weather stations that are predominantly influenced by others based on the strength of their dependencies. Nodes depicted in black and larger sizes denote higher in-degree centrality. Thick edges represent strong dependencies between the nodes. Fig. 2 reveals that weather stations 10, 23, and 24 are notably influenced by other stations, with station 24 being the most significantly affected.

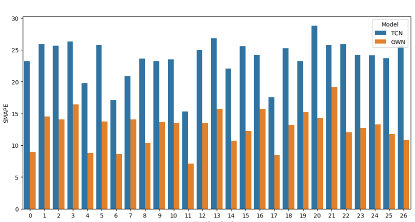

Fig. 1. Comparison of TCN and GWN for the 48-hour prediction horizon

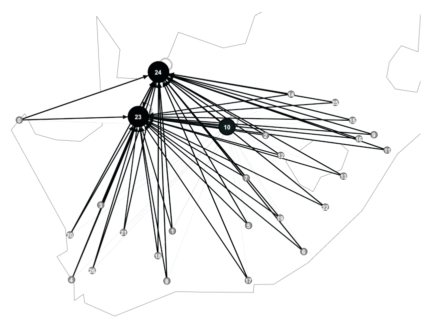

Fig. 2. Visualisation of GWN weighted in-degree centrality measure

4 Discussion

Preliminary results show GWN outperforms TCN, as expected due to STGNNs incorporating spatial information. Visual analysis using weighted in-degree centrality identifies the most influenced weather stations. However, results are limited to a 48-hour prediction horizon. Ongoing research explores longer horizons, compares STGNNs for improved performance, and examines other centrality measures and path analysis to to identify more prominent paths of spatial-temporal dependencies.

References

1. Akram, M., El, C.: Sequence to sequence weather forecasting with long short-term memory recurrent neural networks. International Journal of Computer Applications **143**(11), 7–11 (2016). https://doi.org/10.5120/ijca2016910497
2. Davidson, M., Moodley, D.: St-gnns for weather prediction in south africa. In: Artificial Intelligence Research: Third Southern African Conference, SACAIR 2022, Stellenbosch, South Africa, December 5–9, 2022, Proceedings. pp. 93–107. Springer (2022)
3. Hewage, P., Trovati, M., Pereira, E., Behera, A.: Deep learning-based effective fine-grained weather forecasting model. Pattern Analysis and Applications **24**(1), 343–366 (2020). https://doi.org/10.1007/s10044-020-00898-1
4. Hua, J., Huang, M.L., Huang, W., Zhao, C.: Applying graph centrality metrics in visual analytics of scientific standard datasets. Symmetry **11**(1), 30 (2019)
5. Sahili, Z.A., Awad, M.: Spatio-temporal graph neural networks: A survey. arXiv preprint arXiv:2301.10569 (2023)
6. Shi, X., Chen, Z., Wang, H., Yeung, D.Y., Wong, W.K., Woo, W.c.: Convolutional lstm network: A machine learning approach for precipitation nowcasting. Advances in neural information processing systems **28** (2015)
7. Wilson, T., Tan, P.N., Luo, L.: A low rank weighted graph convolutional approach to weather prediction. 2018 IEEE International Conference on Data Mining (ICDM) (2018). https://doi.org/10.1109/icdm.2018.00078
8. Wu, Z., Pan, S., Long, G., Jiang, J., Zhang, C.: Graph wavenet for deep spatial-temporal graph modeling. Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (2019). https://doi.org/10.24963/ijcai.2019/264

Automatic Grading of First-Year HTML and CSS Assignments using Large Language Models

First Author[1[0000-1111-2222-3333]] and Second Author[2[1111-2222-3333-4444]]

[1] Princeton University, Princeton NJ 08544, USA
[2] Springer Heidelberg, Tiergartenstr. 17, 69121 Heidelberg, Germany
lncs@springer.com

Abstract. The significant increase in university students enrolling for introductory computer science courses makes providing timely and accurate assignment feedback more challenging. The manual grading of assignments, particularly those involving code review, creates a bottleneck that impairs effective student skill development. The study explores how artificial intelligence (AI) and, specifically, large language models (LLMs) can enhance the automated grading of assignments and analyses the scalability and consistency of automated grading.
The first phase of the research focuses on developing an automated grading system. The system extracts critical requirements from a supplied specification document and answer sheet and uses the criteria to grade student submissions.
Advanced AI techniques, such as Reflection, Tool Use, and Multi-Agent Collaboration, are employed to improve feedback, accuracy and relevance.
The second phase compares the effectiveness of local-inference LLMs (e.g., Llama, Gemma) to API-inference models (e.g., ChatGPT). The goal is to identify the most efficient model based on grading time, cost, feedback consistency, and accuracy. Preliminary findings suggest a significant reduction in grading time (up to 97%) and an advantage for API-inference models, as they leverage powerful large-scale infrastructure that users typically do not have access to.
The research highlights AI's potential to improve grading efficiency and offer scalable solutions adaptable to various educational contexts, including coding languages. This system could revolutionise grading at the University of the Free State and beyond, with future work focusing on refining the approach based on user feedback.

Keywords: GPT-3.5 Turbo, GPT-4, LLama3, Gemma-2, large language models, in-context learning, Automated Grading, HTML, CSS, Educational Technology

Algorithm Marginalization: A Problem Beyond Technological Solutions

Bongekile Mkrola[1][0000-0002-6363-7899]

[1] University of Pretoria, Pretoria, South Africa

u18025103@tuks.co.za

Abstract. In this paper I will argue that algorithmic marginalisation within the South African banking sector resulting from biases embedded in Artificial Intelligence (AI) systems utilised in these institutions is not solely an algorithmic problem requiring a technological fix. While the AI technology plays a vital role in bringing these biases to light, at their core, the biases are deeply rooted in social structures that shape the data and the decisions the AI systems rely on. This makes it a necessary condition for the remedy to be implemented to address algorithmic marginalization in banks to examine the social structures which enable the data collected to exhibit patterns which marginalise people. The proposed remedy in this paper is relational ethics by Abeba Birhane to precede the technical fixes. Relational ethics critiques the overreliance on technical solutions to address algorithmic injustice emphasises the importance of putting the individuals affected by these systems at the centre which promotes their humanity rather than treating them as data points.

Keywords: Algorithmic Marginalisation, Banking, Relational Ethics, Bias.

1 Introduction

Marginalization it is often attributed to human acts but seldomly something that can be attributed to AI technology. However, algorithms have infiltrated the social and economic spheres perpetuating the marginalization of some groups of people using biases (Chaka 2022: 84). These biases perpetuate what is termed “digital exclusion, which has much to do with information-poor societies versus information rich societies, or with information have-nots versus information haves” (Chaka 2022: 87). What digital exclusion entails is a system through which “underrepresented users, owing to their socioeconomic, cultural, and historical marginalization, manage to possess only low-level digital devices that are not fitted with user interface designs that can allow them to access essential services online” Chaka 2022: 87). However, digital exclusion is one aspect of a broader issue focused on in this paper which is algorithmic marginalization where biases that are embedded in AI systems perpetuate existing inequalities. In this paper I explore algorithmic marginalization in South African banks and make the argument that it a problem beyond technological solution. Its technological solution must be preceded by the application of relational ethics.

2 Algorithmic Marginalization: A Social Problem

In South Africa, one of the key areas where AI technology is being widely utilized is in the banking sector. AI-powered systems are being developed for the banking sector to bring about advanced functionality, develop innovative products, and increase efficiency (Sheth, Jain, Roy, and Chakraborty 2022). Additional applications of AI in banking include chatbots, which provide support to customers around the clock via a human-like interface; fraud prevention and detection; customer management; predictive analytics; flagging of questionable documents via AI-enabled ID verification; credit risk management; and regulatory reporting (Malinga 2024). Although these innovations are developed with the aim of making people's lives easier, they do not come without disadvantages. One grave concern they come with, is the issue of structural bias in data sets. This type of bias is related to identity prejudice and stereotyping (Kartal 2022) and might cause the outcomes generated by AI systems to discriminate against certain social groups. For example, consider the incident of a Black female entrepreneur who was denied a loan because her skin color made her to be categorized as a risky client based on the data the relevant algorithm was trained on (Moosajee 2019). This incident is an example of what is termed algorithmic marginalization which occurs when biased algorithms systematically disadvantage certain groups, reinforcing and perpetuating existing societal inequalities (Chaka 2022). What happens is that the loan history and demographic data of applicants whom loan officers accept, or reject is used to 'train' AI to determine new loan applications. Although the race and gender information may not explicitly be captured, studies have shown that AI can use location data and habits to predict race and gender and inadvertently discriminate against applicants (Moosajee 2019). This shows that, in banking today, both face-to-face and algorithmic practices exhibit significant discrimination (Pérez, 2022).

This discrimination is an illumination of a deeply seated social issue in South African societies originating from the apartheid era. The discrimination which results in marginalization is social in that it reinforces biases and promotes the exclusion of certain groups of people which mirrors the institutions which perpetuate such an injustice. Given its social nature, my argument is that remedying it using a technological fix is inadequate and reasons to support argument will be discussed in the section below.

3 Challenges Associated with Technological Solutions

The automation of issues which are social in nature is a complex process as it entails the responsibility of making moral and ethical decisions not just technical ones. Utilizing a technical fix to address social issues is a short-term solution that does not address the root issue but deals with it at the surface level. Birhane asserts that,

> "The mathematization and formalization of social issues brings with it a veneer of objectivity and positions its operations as value-free, neutral, and amoral. The intrinsically political tasks of categorizing and predicting things such as ''acceptable'' behavior, ''ill'' health, and ''normal'' body type then pass as apolitical technical sorting and categorizing tasks. Unjust and harmful outcomes, as a result, are treated as side effects that can be treated with technical solutions such as ''debiasing'' datasets rather than problems that have deep roots in the mathematization of ambiguous and contingent issues, historical inequalities, and asymmetrical power hierarchies or unexamined problematic assumptions that infiltrate data practices." (Birhane 2021: 2).

This highlights the inadequacy of technical fixes to address deep seated social issues. When we use technical fixes to address social problems, harmful outcomes are then misinterpreted as mere technical glitches that can be solved with fixes like "debiasing" datasets, rather than as structural problems arising from the flawed application of mathematical models to complex, contingent social realities (Birhane 2021). Within the South African context, historically marginalized communities, particularly Black South Africans have faced structural barriers to accessing financial services, resulting in limited credit histories and financial exclusion (Moosajee 2019). This data is fed back into the AI models which creates a feedback loop that perpetuates financial inequality illuminating the social reality of the data sets. This results in the communities that have historically been denied access to banking services or fair credit terms to be further marginalized by these algorithmic systems. While these groups of people are overrepresented in datasets related to financial risk, they are underrepresented in data concerning successful financial behavior or credit growth because of past data sets. What becomes a significant challenge to remedy this injustice even if an algorithm is reprogrammed with the aim of ensuring fairness, a deeper issue remains: "we are left with the deeper philosophical and political issue of whether neutrality constitutes fairness in background conditions of pervasive inequality and structural injustice" (Zimmermann et al., 2023). When we examine neutral solutions such as the ones we try to program into algorithms, these solutions "further entrench existing injustices" (Zimmermann et al., 2023). This is because, even if algorithms are neutral, the data that the algorithms are trained on is riddled with prejudice (Zimmermann et al., 2023). This raises the question of whether a better algorithm is the solution to these social problems, as the ineffectiveness of such technical fixes brings to light critical ethical concerns. In the following section I will present a remedy that I propose should precede the technical fixes.

4 Relational Ethics

In this section explore relational ethics proposed by Abeba Birhane as a solution that should precede technical fixes when it comes to remedying algorithmic marginalization. It has been noted that "the growing body of work exposing algorithmic injustice has indeed brought forth increased awareness of these problems, subsequently spurring the development of various techniques and tactics to mitigate bias, discrimination, and harms. However, many of the ''solutions'' put forward (1) revolve around technical fixes and (2) do not center individuals and communities that are disproportionally impacted" (Birhane 2021: 2). My contention is that relation ethics is a better method to precede technical fixes in addressing algorithmic marginalization because it firstly, critiques the overreliance on technical solutions to address algorithmic injustice. Here there is an emphasis that many "technical solutions" fail to center the individuals and communities most impacted by these technologies. Secondly, relational ethics demands that we prioritize the lived experiences and needs of marginalized communities. This includes acknowledging historical injustices and the ongoing impact of AI systems on these vulnerable groups. Thirdly, the framework compels us to re-examine our assumptions, question hierarchical power structures, and consider the broader context in which algorithmic systems are deployed. It stresses the importance of understanding the interconnected background that shapes these technologies. Fourthly, the framework promotes a shift from focusing solely on datasets to considering the contexts referred to as data setting as this is vital in highlighting that behind the data points, there are people subjected to algorithmic marginalization. Lastly, the framework of relational ethics challenges the notion that bias can be eliminated from datasets. It urges us to look deeper into the structural issues, historical contexts, and power dynamics that perpetuate inequalities and try to come up with a remedy to address these before considering technical fixes.

5 Conclusion

In this paper, I explored how algorithmic marginalization in the South African banking sector is not merely a technological issue but rather a reflection of the social structures that shape both data and AI systems. Therefore, proposing that a solution like relational ethics, which emphasizes human relationships and interconnectedness, is necessary to address these structural injustices and it should precede the technical fixes to address the root problem.

References

1. Chaka, C., 2022. Digital marginalization, data marginalization, and algorithmic exclusions: A critical southern decolonial approach to datafication, algorithms, and digital citizenship from the Souths. Journal of e-Learning and Knowledge Society, 18(3), pp.83-95.
2. Kartal, E., 2022. A Comprehensive Study on Bias in Artificial Intelligence Systems: Biased or Unbiased AI, That's the Question! International Journal of Intelligent Information Technologies (IJIIT), 18(1), pp.1-23.
3. Malinga, S. (2024) Big-four Banks take lead in SA's GenAI, AI deployments, ITWeb. Available at: https://www.itweb.co.za/article/big-four-banks-take-lead-in-sas-genai-ai-deployments/G98YdqLGK9pMX2PD.
4. Moosajee, N. (2019). Fix AI's Racist, Sexist Bias. The Mail & Guardian. Available at: https://mg.co.za/article/2019-03-14-fix-ais-racist-sexist-bias/. (Accessed: 17 July 2024).

5. Pérez, Á., 2022. Bias in Artificial Intelligence Models in Financial Services. Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society. https://doi.org/10.1145/3514094.3539561.
6. Sheth, J., Jain, V., Roy, G., & Chakraborty, A. (2022). AI-driven Banking Services: The next Frontier for a Personalized Experience in the Emerging Market. The International Journal of bank Marketing, 40(6), 1248-1271. https://doi.org/10.1108/ijbm-09-2021-0449
7. Zimmermann , A., Di Rosa , E. and Kim, H. (2023) Technology can't fix algorithmic injustice, Boston Review. Available at: https://www.bostonreview.net/articles/annette-zimmermann-algorithmic-political/ (Accessed: 10 September 2024).

Implications of Privacy and Consent on Data-Driven Decision-Making Adoption in Higher Education

First Author[1[0000-1111-2222-3333]] and Second Author[2[1111-2222-3333-4444]]

[1] Princeton University, Princeton NJ 08544, USA
[2] Springer Heidelberg, Tiergartenstr. 17, 69121 Heidelberg, Germany
lncs@springer.com

Abstract. This paper explores privacy and consent challenges in data-driven decision-making in higher education, focusing on student data. Semi-structured interviews were conducted with 10 participants and thematically analysed. Findings reveal that students do not willingly consent to the use of their data by institutions, as they argue that there is no transparency in the processing and use of their data. It was further found that students believe that giving consent to their data violates their privacy, but they cannot opt-out due to fear of being denied essential services. The findings are useful for policymaking in higher education, highlighting the need to address privacy and consent issues so that data-driven decision-making is ethically done. Thus, the study contributes to theory and practice, as it also adds valuable information that most institutions take for granted, provoking debates that lead to transformation in how data is collected, stored and processed within higher education institutions.

Keywords: Data-driven decision-making, data analytics, higher education, informed consent, privacy.

1 Introduction and Background

The use of data for decision support is on the rise in higher education institutions globally, with many institutions increasingly relying on data-driven decision-making to enhance student outcomes, optimize resources, and improve institutional effectiveness (Kovač & Oreški, 2018; Nouri et al., 2019; Gil et al., 2021). Data-driven decision-making in higher education offers many benefits to students and the institution at large. However, there is a plethora of ethical considerations and challenges to be addressed to harness the potential benefits of data-driven decision-making. For example, there are ongoing debates on issues of privacy and consent to data use (Jones, 2019; Khan, Usman & Moinuddin, 2024). This study aims to explore issues of data privacy and student consent to data use that arise from data-driven decision-making in higher education and suggests ways of overcoming them. Thus, the findings of the study contribute to enhancing the adoption and ethical use of data-driven decision-making in higher education. A case study of a South African higher education institution is used, adopting an interpretive approach.

2 Methodology

This paper presents the findings of a qualitative study that adopted an interpretive approach. Ethical issues on data privacy, ethics and consent are subjective, thus the choice of interpretivism (Saunders, Lewis & Thornhill, 2009; Bryman, 2012). Data was collected through semi-structured interviews with 8 students and 2 data specialists, and a literature survey. Literature was systemically reviewed from the results of searches using the keywords that included "data-driven decision-making," "higher education," "ethics," "data privacy," "data ethics", "learning analytics" and "informed consent". Articles were selected based on their relevance to ethical concerns and the use of empirical evidence to support their claims. The themes from the articles were merged with the ones from the interviews.

3 Results and discussions

The findings show that students are concerned with their privacy on the use of their data for decision-making. In addition, they claim that the consents they sign are not informed, as they are not fully aware of how and what their data will be used. Unfortunately, they feel they have no choice but to consent, or risk being denied access to various essential services necessary for student success.

3.1 Privacy Concerns

Institutions collect vast amounts of student data that is supposed to advise operations and benefit students. Unfortunately, this data is prone to a potential violation of student privacy. The use of student data has the potential to breach students' privacy. According to Tsai & Gasevic, (2017) and Selowa, Ilorah & Mokwena, (2022), the use of data analytics in higher education involves the collection of large volumes of personal information, including academic records, social interactions, and even biometric data. This raises questions about security measures to safeguard the data, as well as who has access to this data. Tsai & Gasevic, (2017) this can be overcome by establishing transparent, clear data collection, access and storage guidelines.

3.2 Informed Consent

Informed consent is a foundational ethical principle in data collection, but it faces significant challenges in the context of data-driven decision-making in higher education. Diverse types of personal data are collected, analysed and used for various uses including enhancing student experiences and institutional performance. According to Selowa, Ilorah & Mokwena, (2022), the processes and technologies involved in data collection and analysis are complex, and students may not fully understand what they consent to when agreeing to use their data. This undermines the idea of informed consent, as it is impossible to make informed decisions about how their data will be used without a full understanding of the processes.

Furthermore, institutions are not always transparent about how they use data (Lepri et al., 2018; Muharlisiani et al., 2023). As such, students may not know how their data is being used, and who has access to it. Jones, (2019) argues that students cannot truly give informed consent without full disclosure, making the issue of transparency an ethical dilemma in data-driven decision-making in higher education. Institutions should be transparent in their data processes so that data subjects give true consent.

Informed consent should include room for opting out (Roberts et al., 2016; Jones, 2019). However, in higher education, there is little to no option for opting. This is because institutional data collection is often automatic, and embedded in online learning platforms, student information systems, and other digital tools that students use regularly (Wong, 2017). Thus, opting out significantly impacts one's ability to participate in academic activities. This makes students feel obliged to consent to avoid negative consequences or restricted access to essential academic and support services. In addition, informed consent is considered a once-off issue, yet data collection and processing are ongoing throughout a student's lifespan at the institution(Jones, 2019). During this period, the uses and processing of data may change, thus invalidating the previously given consent.

Students highlighted concerns regarding how algorithms and data analytics tools work to reach decisions about their academic performance and support. According to (Jones, 2019), algorithms often make decisions based on patterns in student data that students themselves do not fully understand. This makes consenting challenging, as students cannot fully assess the risks and benefits they agree to.

4 Implications and Conclusion

Ethical issues around data privacy and informed consent have significant implications for higher education institutions. Higher education institutions stand to benefit from adopting data-driven decision-making approaches. However, without addressing crucial privacy and ethical issues, the adoption of data-driven decision-making will remain a challenge. The findings of this study are crucial for policymakers and institutional leadership to revisit their data privacy policies, ensuring that they abide by ethical guidelines and legislation. Furthermore, data subjects must be informed of how their data is used, benefits and risks at every stage

References

1. Bryman, A. 2012. Social Research Methods. Fourth ed. Oxford University Press.
2. Gil, P.D., da Cruz Martins, S., Moro, S. & Costa, J.M. 2021. A data-driven approach to predict first-year students' academic success in higher education institutions. Education and Information Technologies. 26(2):2165–2190. DOI: 10.1007/S10639-020-10346-6/METRICS.

3. Jones, K.M.L. 2019. Learning analytics and higher education: a proposed model for establishing informed consent mechanisms to promote student privacy and autonomy. International Journal of Educational Technology in Higher Education. 16(1). DOI: 10.1186/S41239-019-0155-0.
4. Khan, R., Usman, M. & Moinuddin, M. 2024. From Raw Data to Actionable Insights: Navigating the World of Data Analytics. International Journal of Advanced Engineering Technologies and Innovations. 1(4):142–166. Available: https://ijaeti.com/index.php/Journal/article/view/267.
5. Kovač, R. & Oreški, D. 2018. Educational Data Driven Decision Making: Early Identification of Students at Risk by Means of Machine Learning. In The Central European Conference on Information and Intelligent systems. 231–237.
6. Lepri, B., Oliver, N., Letouzé, E., Pentland, A. & Vinck, P. 2018. Fair, Transparent, and Accountable Algorithmic Decision-making Processes: The Premise, the Proposed Solutions, and the Open Challenges. Philosophy and Technology. 31(4):611–627. DOI: 10.1007/s13347-017-0279-x.
7. Muharlisiani, L.T., Mulawarman, W.G., Suwarni, S., S, U., Hutahaean, B. & Rahim, R. 2023. Designing a Decision Support System for Educational Resource Allocation. AL-ISHLAH: Jurnal Pendidikan. 15(3):4216–4225. DOI: 10.35445/ALISHLAH.V15I3.4120.
8. Nouri, J., Ebner, M., Ifenthaler, D., Saqr, M., Malmberg, J., Khalil, M., Bruun, J., Viberg, O., et al. 2019. Efforts in Europe for Data-Driven Improvement of Education A Review of Learning Analytics Research in Seven Countries. International Journal of Learning Analytics and Artificial Intelligence for Education (iJAI). 1(1):8–27.
9. Roberts, L.D., Howell, J.A., Seaman, K. & Gibson, D.C. 2016. Student attitudes toward learning analytics in higher education: “The fitbit version of the learning world”. Frontiers in Psychology. 7(DEC). DOI: 10.3389/FPSYG.2016.01959/FULL.
10. Saunders, M., Lewis, P. & Thornhill, A. 2009. Research Methods For Business students. Fisth ed. Prentice Hall.
11. Selowa, K.T., Ilorah, A.I. & Mokwena, S.N. 2022. Using Big Data analytics tool to influence decision-making in higher education: A case of South African Technical and Vocational Education and Training colleges. SA Journal of Information Management. 24(1):1–8. DOI: 10.4102/sajim.v24i1.1489.
12. Tsai, Y.-S. & Gasevic, D. 2017. Learning analytics in higher education---challenges and policies: a review of eight learning analytics policies. The seventh international learning analytics &. (March, 13):233–242. DOI: 10.1145/3027385.3027400.
13. Wong, B.T.M. 2017. Learning analytics in higher education: an analysis of case studies. Asian Association of Open Universities Journal. 12(1):21–40. DOI: 10.1108/AAOUJ-01-2017-0009.

EXTENDED RESEARCH ABSTRACT

RESEARCH TOPIC: **CAN ARTIFICIAL INTELLIGENCE (AI) BE GOOD DOCTORS? A PHILOSOPHICAL ANALYSIS OF MEDICAL ARTIFICIAL INTELLIGENCE**

Recent discussion on Artificial Intelligence (AI) excelling as doctors has gained significant attention due to advancements in the AI industry and its application in medicine. This progress is demonstrated through AI's adoption of a mathematical algorithmic process called Machine Learning. Although there are several interpretations of this process, machine learning is simply the technique by which computer algorithms gain knowledge from the huge data at their disposal, and then use that knowledge to generate predictions as outcome [1, 2]. Thus, machine learning is a subset of AI that can perform tasks based on the data it has been exposed to and integrated.

The advancement of Artificial Intelligence (AI) and its capacity to learn from extensive datasets, particularly in medicine, has initiated discussions about the potential replacement of human medical clinicians in all medical domains due to their remarkable efficiency. For instance, Silicon Valley entrepreneur and investor Vinod Khosla predicted the obsolescence of the role of radiologists within five years[3]. Furthers this claim by stating that the advent of "technological singularity" (a hypothetical point in the future when AI surpasses human intelligence) will inevitably lead to the replacement of medical practitioners by AI systems[4].

While optimism exists regarding the potential of AI to replace clinicians, opposing viewpoints highlight various concerns. Roselin McDougall[5] raises a significant issue, contending that AI systems suggesting treatment options may jeopardize shared decision-making, as the ranking of treatments lacks the influence of individual patient values. Such might stand as an impediment to the trust that ought to exist between patients and doctors. As Hatherley[6] suggests, trust is essential in medicine, and there are doubts about relying on AI as they are not deemed an "appropriate object of trust." Furthermore, Duran and Jongsman[7] expound on this lack of trust in AIs, by drawing attention to the opaque nature of AI algorithms in their tasks, questioning the validity of trusting diagnoses derived from algorithms when the process remains undisclosed.

Now, although discourses on medical AI have been occupied with debates about the possibility of it replacing clinicians, none have thoroughly explored the question if they can be good doctors. If AIs through machine learning can perform clinician’s tasks like medical diagnosis, prognosis, and treatment recommendation reliably and effectively, can we regard them as good doctors? **I argue that given how AIs can carry out clinical tasks such as diagnosis, prognosis, treatment recommendation, and their ability to assimilate a huge dataset and self- improve, they should be considered good doctors.**

To sustain my argument, I first consider and examine what it means to be a good doctor. I will do this by drawing from various existing literature, specifically those contained in the evidence-based medicine (EBM) versus the patient-centred care debate (PCC). The rationale for my focus on the foregoing stated debate is because the two approaches are the most prominent in the literature on how medicine is to be carried out. More so, they both provide guidelines to clinicians on how to carry out medical practices effectively. I will then investigate if (the features of) what it means to be a good doctor as provided by either or any of the approaches I will settle for, can be accommodated by AI (algorithms). I will also consider some of the alleged limitations that stand in the way of medical AIs being fully accepted in the medical fares (such as their incapacity for shared-decision, algorithmic biases etc). I will then suggest that medical AIs can integrate these features, especially those features that are directly and indirectly tied to various human values. Furthermore, I will use my analysis of how AIs are able to assimilate huge medical datasets (via machine learning), which has consequently prompted their effectiveness in diagnosis, to show how they can integrate human values in the form of data in carrying out their clinical tasks.

The relevance of my undertaking bears on first, the ethical implications it warrants, as examining AI's involvement in healthcare forces one to think about the moral ramifications of technology making crucial choices in medical settings. Talks about the proper usage of AI can benefit from this analysis. Second, assessing AI's potential as a medical professional is essential to determining how it will affect patient care and safety. Verifying the precision and dependability of AI-based diagnostics and prognosis is crucial to guaranteeing good health results. Also, a discourse on AI doctors can help advance conversations about how to control the application of AIs in healthcare to guarantee standards compliance and accountability, thus establishing legal and regulatory framework for its adoption. Finally, the query encourages a philosophical investigation of the interplay between human competence and technology in healthcare. This approach can add to more general conversations concerning the incorporation of technology into fields that have historically been dominated by humans.

Reference List

1. Hamet, P. and J. Tremblay, *Artificial intelligence in medicine.* metabolism, 2017. **69**: p. S36-S40.
2. Nagy, M., N. Radakovich, and A. Nazha, *Machine learning in oncology: what should clinicians know?* JCO Clinical Cancer Informatics, 2020. **4**: p. 799-810.
3. Aminololama-Shakeri, S. and J.E. López, *The doctor-patient relationship with artificial intelligence.* American Journal of Roentgenology, 2019. **212**(2): p. 308-310.
4. Shuaib, A., H. Arian, and A. Shuaib, *The increasing role of artificial intelligence in health care: will robots replace doctors in the future?* International journal of general medicine, 2020: p. 891-896.
5. McDougall, R.J., *Computer knows best? The need for value-flexibility in medical AI.* J Med Ethics, 2019. **45**(3): p. 156-160.
6. Hatherley, J.J., *Limits of trust in medical AI.* Journal of medical ethics, 2020. **46**(7): p. 478-481.
7. Durán, J.M. and K.R. Jongsma, *Who is afraid of black box algorithms? On the epistemological and ethical basis of trust in medical AI.* Journal of Medical Ethics, 2021. **47**(5): p. 329-335.

Temporal Hierarchical Time Series Modelling: A Comparative Analysis of Aggregation Techniques and Model Efficiency

Author, Callyn Blayne Barath, Author, Edgar Jembere
University of Kwa-Zulu Natal
220010761@stu.ukzn.ac.za
jemberee@ukzn.ac.za

1. INTRODUCTION

For retail operations, accurate sales forecasting plays a pivotal role in shaping decisions regarding inventory management, supply chains, and overall business strategy.The ability to reliably forecast sales is critical to the working of retail operations, affecting inventory handling, supply chain planning, as well as general business direction [1]. However, traditional time series forecasting methods tend to ignore the hierarchical nature of sales data and attempt to model the data at the individual aggregation level (e.g., weekly, monthly, quarterly). In this study, we applied temporal hierarchical time series models to improve forecasting accuracy and reliability by aggregating and disaggregating information across varying temporal levels [2].

Our work aimed to evaluate how effective are temporal hierarchical aggregation and reconciliation methods on improving forecasting accuracy. The performance of hierarchical models was compared with the performance of traditional individual time series models at each aggregation level. The study also investigates whether machine learning models outperform classical statistical models in this hierarchical framework.

2. METHODS

This study employs a dataset from a major South African retailer and applies temporal hierarchical time series models using various aggregation levels—annual, semi-annual, quarterly, bi-monthly, and monthly [3]. We use classical statistical models (e.g., AutoARIMA, Seasonal Naive, Holt-Winters) and machine learning models (e.g., Random Forest Regression, XGBRegressor, MLP Regressor) to generate base forecasts. These forecasts are then reconciled across the temporal hierarchy using three reconciliation methods: Bottom-Up, Top-Down, and Middle-Out. A sparse matrix framework is employed to align base forecasts from different time granularities and ensure consistency across temporal levels [3].

To evaluate the models' performance, we employ standard forecasting accuracy metrics, including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R-squared (R^2). Additionally, we apply a square root transformation to stabilize the variance and reduce the influence of outliers in the data before model fitting [4]. This transformation is reversed after predictions are made, allowing for a meaningful comparison of the forecasts against actual sales figures.

3. RESULTS AND DISCUSSION

The main finding wastemporal hierarchical models doing better in improving forecast accuracy than standard individual time series models. Models were aggregated and disaggregated across different temporal levels to better capture critical seasonal patterns and trends in retail sales data. Especially at more granular levels of what, say, monthly and bi-monthly, Mean Absolute Error (MAE), Root

Mean Square Error (RMSE), and R-squared (R^2) error measures demonstrated notable reductions in what was forecasted.

In annual and semiannual aggregation levels, the models tended to do worse, as the number of observant observations at those levels is limited. RMSE and MAE values increase as the number of data points available to models to generalize to decreases. Hierarchical models performed better than nonhierarchical ones, showing that temporal hierarchical modeling is very helpful to preserve forecast accuracy under different scales of aggregation.

We show that for each level of aggregation, traditional time series models time series models are outperformed by temporal hierarchical models employing reconciliation methods such as Bottom-Up, Top-Down and Middle-Out. The models' power to explain variance in the data is reflected by the R^2 values (the improvements were particularly evident in the R^2 values).

In addition we aimed to compare the performance of machine learning models (RandomForestRegressor, XGBRegressor, MLP Regressor) with classical statistical models (AutoARIMA, Holt-Winters, Seasonal Naive). RMSE and MAE values of XGBRegressor and RandomForestRegressor turned out to be substantially lower, especially with more granular aggregation levels (monthly and bi-monthly). Even further, R^2 values helped to further emphasize the superiority of machine learning models, with XGBRegressor consistently out performing statistical models like AutoARIMA in explaining the variance.

The results also show that sample size and data scale can affect the performance of forecast models. Errors (RMSE and MAE) increased at higher aggregation (annual and semi-annual) due to the reduced ability of models to generalize at higher aggregation levels. Most pronounced on these higher error metrics were classical models, which suffer from limited data.

Secondly, the evaluation metrics for RMSE and MAE were affected by data scale. Naturally large sales figures inflate these metrics, making it hard to compare models without looking at scale. The variance in the data was stabilized when applied through a square root transformation before model training, reducing the variance in the data and making these performance metrics more comparable amongst models.

4. CONCLUSION

This study details how temporal hierarchical time series models achieve significant improvements in forecasting accuracy over traditional individual time series models. Overall, the hierarchical approach, especially when using Middle Out reconciliation, consistently outperformed middle out reconciliation across all metrics, including MAE, RMSE and R squared, such that the improvement was more significant at finer temporal levels such as monthly and bi monthly. The fact that temporal aggregation and disaggregation can effectively capture both the trends and seasonality in retail sales data is illustrated here.

Additionally, the results indicate that classical statistical models (e.g., Linear Regression, Polynomial Regression) perform poorly in general, and that machine learning models (e.g., XGBRegressor, RandomForestRegressor) generally outperform classical statistical models, especially in situations where relationships in the data are complex and non-linear. Hierarchical reconciliation added a small amount of extra value in addition to already better forecasting accuracy, and demonstrated its use when reconciling forecasts in complex retail forecasting scenarios. Future work will continue to investigate more advanced machine learning methods and hybrid models to increase forecast reliability and operable decision making in the retail industry.

5. REFERENCES

[1] F. Theodosiou, and N. Kourentzes, “Forecasting with deep temporal hierarchies,” *Available at SSRN 3918315,* 2021.

[2] S. S. Rangapuram, S. Kapoor, R. S. Nirwan, P. Mercado, T. Januschowski, Y. Wang, and M. Bohlke-Schneider, "Coherent probabilistic forecasting of temporal hierarchies." pp. 9362-9376.

[3] G. Athanasopoulos, R. J. Hyndman, N. Kourentzes, and F. Petropoulos, “Forecasting with temporal hierarchies,” *European Journal of Operational Research,* vol. 262, no. 1, pp. 60-74, 2017.

[4] T. O. Hodson, “Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not,” *Geoscientific Model Development Discussions,* vol. 2022, pp. 1-10, 2022.

A Comparative Study of Contemporary Cross Sectional Hierarchical Time Series Reconciliation Techniques

Ahana Maharaj and Edgar Jembere

Department of Computer Science, University of KwaZulu-Natal, Westville, Durban, 3629, KwaZulu-Natal, South Africa.

Contributing authors: maharaj.ahana1@gmail.com; jemberee@ukzn.ac.za;

1 Introduction

Cross sectional hierarchical time series models consist of multivariate time series that fit within a hierarchical (tree) structure, where time series lower in the hierarchy must aggregate to time series higher in the hierarchy. When this fundamental condition is met, the hierarchy is referred to as being coherent. Forecasting these hierarchical time series structures is critical as they are able to aid decision making at many levels of granularity in areas such as health-care, urban planning and retail.

Producing accurate and coherent forecasts for these hierarchical structures is a challenge. Therefore, a novel two stage approach was created, where time series in the hierarchy are first independently forecasted, and then combined and modified in the reconciliation step to generate optimal forecasts. These reconciled forecasts guarantee coherency and improve the accuracy of the hierarchical time series predictions.

An issue, however, is that these models only produce mean point forecasts, and thus are not able to quantify uncertainty, which can be important when taking risks in decision-making. Therefore, there is also a movement towards probabilistic reconciliation methods which are able to determine uncertainty.

This evolution of reconciliation methods is indicative of growing academic interest in the area. Despite this trend, there has been little discussion on the comparison between the various techniques and how well they perform. In fact, multiple papers have results that seemingly contradict another. Therefore, in this paper, the aim is to find which contemporary hierarchical time series reconciliation technique results in the optimal predictions at all levels of granularity in cross sectional hierarchical time series modelling. We hope that this also provides a guideline for future comparative analysis of reconciliation methods.

2 Methodology

The code was implemented using Python on the Google Colab platform. A variety of libraries from NITXLA, including the HierarchicalForecast and StatsForecast were utilised.

Two different publicly available benchmark datasets were utilised. The first hierarchical time series dataset is the Quarterly Australian Tourism Dataset which records quarterly domestic tourism demand in Australia. It consists of 5 levels in the hierarchy with 80 time steps and 85 total time series in the hierarchy. On the other hand, the Wiki2 records the daily Wikipedia article views. It provides data for 80 time steps, with 5 different levels in the hierarchy, which is made up of 58 individual time series. Since these hierarchical datasets have a different structure and varying, complex dependencies within their levels of granularity, the results from the evaluations will be more consistent.

Similarly two different models for the independent forecasting were carefully chosen. The first individual time series forecasting model selected was AutoARIMA, which can automatically select the optimal parameters for the ARIMA (Autoregressive Integrated Moving Average) model. The second model selected was XGBRegressor, a tree-based supervised machine learning method. Again, by selecting two different models, we are assured that results are consistent.

In the comparative analysis, the contemporary point reconciliation methods MinT (Minimum Trace) and ERM (Empirical Risk Minimisation) are evaluated using the MASE (Mean Absolute Scaled Error) and MAPE (Mean Absolute Percentage Error) evaluation metrics due to their scale independence. MAPE is the calculated mean for the of the percentage error, which provides good interpretibility but can give erratic results when observations are near zero. On the other hand MASE, calculates the mean of the scaled error and is considered to be less effected by outliers. For the probabilistic reconciliation, PERMBU (Pooled Error Regression for MinT with Bottom-Up) and Bootstrapping were the methods chosen for comparative study. CRPS (Continuous Ranked Probability Score) which was selected over other probabilistic evaluation metrics like log score, due to its independence to the underlying data distribution.

3 Results

The code was implemented using Python on the Google Colab platform. A variety of libraries from NITXLA were utilised.

A survey of the results show that the MinT methods consistently performed better than ERM reconciliation. Interestingly, MinT did sometimes show more erratic results with the XGBRegressor forecasts, which suggests that maybe using ERM when forecasting with XGBRegressor may be a better option. With regards to probabilistic forecasting, the PERMBU method was consistently outperformed by the bootstrapping method.

4 References

“Time Series Forecasting: Definition, Applications, and Examples,” Tableau, 2024. [Online]. Available: https://www.tableau.com/learn/articles/time-series-forecasting

G. Athanasopoulos, R. J. Hyndman, N. Kourentzes and A. Panagiotelis, “Forecast reconciliation: A review,” International Journal of Forecasting, 2023

S. B. Taieb, J. W. Taylor and R. J. Hyndman, “Coherent Probabilistic Forecasts for Hierarchical Time Series,” in 34th International Conference on Machine Learning, Sydney, 2017.

B. Basener, “Machine Learning 4.2 - Bootstrapping,” 12 December 2020. [Online]. Available: https://www.youtube.com/watch?v=WAFTWyu-hn0.

D. Girolimetto, G. Athanasopoulos, T. D. Fonzo and R. J. Hyndman, “Cross-temporal probabilistic forecast reconciliation: Methodological and practical issues,” International Journal of Forecasting, pp. 1134-1151, 2023.

G. Mélard, J. M. Pasteels, ”Automatic ARIMA modeling including interventions, using time series expert software”, International Journal of Forecasting, pp. 497-508, 2000.

S. L. Wickramasuriya, G. Athanasopoulos and R. J. Hyndman, “Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization,” Journal of the American Statistical Association, 2018

X. Han, S. Dasgupta and J. Ghosh, “Simultaneously Reconciled Quantile Forecasting of Hierarchically,” in The 24th International Conference on Artificial Intelligence and Statistics (AISTATS), San Diego, 2021

R. J. Hyndman, “Professor Rob J Hyndman: Ten years of forecast reconciliation,” International Institute of Forecasters, 28 October 2020. [Online]. Available: https://www.youtube.com/watch?v=5jB09RsKOct=2345s.v=5jB09R-sKOct=2822s

D. Girolimetto, G. Athanasopoulos, T. D. Fonzo and R. J. Hyndman, “Cross-temporal probabilistic forecast reconciliation: Methodological and practical issues,” International Journal of Forecasting, pp. 1134-1151, 2023

G. Athanasopoulos, R. A. Ahmed and R. J. Hyndman, “Hierarchical Forecasts for Australian Domestic Tourism,” International Journal of Forecasting 25, pp. 146-166, 2009.

R. J. Hyndman and A. B. Koehler, “Another Look at Measures of Forecast Accuracy,” International Journal of Forecasting, p. 679– 688, 2006

Propolingo: a language learning application for propositional logic

Bonga Baleni

University of the Western Cape, Bellville, South Africa
4020552@myuwc.ac.za

Inspired by Duolingo's gamified language learning approach, Propolingo is gaming application that is my current research project,using similar techniques from Duolingo to simplify propositional logic for students, offering interactive tasks and instant feedback https://www.duolingo.com/. The capacity for learning is intrinsic to our nature as human beings. The field of artificial intelligence (AI) strives to replicate these human cognitive processes in machines,for example to develop computer systems capable of intelligent decision-making and learning [3].The field in computer science that is concerned with how information is represented and how to derive new information is called knowledge representation and Reasoning (KRR). The key aspects in knowledge representation include logic-based representation, semantic networks, frames, production rules, and ontologies.Various logic-based representations, or formalisms exist within KRR. Examples are classical propositional logic and higher order logics such as description logic and predicate logic.In this text, we will be focusing on propositional logic.According to Ben Ari, propositional logic is a formal system in which formulas represent statements that are either true or false, and logical connectives are used to build more complex statements from simpler ones [1]. A complex propositional statement can be made using logical connectives between atomic propositions, and such complex statements are often referred to as formulas in propositional logic. The logical connectives are disjunction (OR, $\vee$), conjunction (logical and, $\wedge$), negation (logical not, $\neg$), implication (conditional if, $\rightarrow$), and bi-conditional (if and only if, $\leftrightarrow$) [1]. An example of a formula in propositionsal logic would be R $\rightarrow$ Q which can be read as "If it is cloudy , then it is raining".A formula is assigned a truth value by considering the truth values of the the atoms and applying logic connectives, when different combinations of truth values are assigned to atoms we get different interpretations of the formula.We can understand interpretations to mean the different rows in a truth table for any propositional formula. Each row is a unique assignment of true and false to each atom in the formula. This gives rise to a single interpretation, or possible world. When one formula follows logically, or can be derived from another formula using logical rules, we call this this relationship entailment. If $R \models Q$, then R being true implies Q is true. If there exists an instance where R is true and Q is false, then $R \not\models Q$.We have established that propositional logic is a useful representation language for two-valued statements, and it can be used to derive higher-order logics.However, one of the challenges in the South African education system is the lack of adequate exposure to this field at undergraduate level.For instance, a study conducted by Maharaj [2] found that 68%

of students at a South African university struggled with basic concepts of formal logic. This paper introduces *Propolingo* as a gamified application designed to teach students struggling with foundational concepts in propositional logic. This application aims to improve upon existing tools, such as *Logicly* and *Tarski's World*, by adapting the content to fit the South African university curriculum. Unlike previous applications, *Propolingo* focuses on addressing specific gaps in students' understanding by incorporating curriculum-relevant examples, interactive problem-solving tasks that align with the logical reasoning required in South African computer science and AI courses. Additionally, it emphasises gradual progression in complexity, from basic atoms to more advanced topics such as entailment, ensuring students build a solid foundation in formal reasoning.Among others, the Propolingo application involves different functional requirements that allow students to be engaged and committed to learning propositional logic. The application offers interactive exercises with immediate feedback that enables users to learn through real-time corrections. Users will have options where a user selects particular modules or aspects they want to engage in propositional logic. It also makes the learning path customizable; the difficulty can increase as the student progresses. Moreover, it also provides users with the ability to track their progress and find weak points, while the application offers useful hints for basic controls that make learning easier. The development of Propolingo follows a user-centered design, ensuring the application addresses students' specific learning needs. The lifecycle involves iterative design and testing phases to refine the user experience.Propolingo includes interactive exercises that challenge students to apply propositional logic concepts. Feedback mechanisms provide immediate responses to the user's actions, helping them correct mistakes in real-time. The adaptive learning path ensures that users can progress from basic concepts to more advanced topics at their own pace. It also has incentives such as points in order to keep the student engaged wanting to achieve more.Propolingo offers big advantages over traditional tools through its gamification, the interactive exercise set, and personalised learning, making it far more accessible and effective for students. The content for this prototype is geared toward an introductory course that teaches the language, interpretations and models and entailment, but no reasoning algorithms. Future development may expand the content, introduce adaptive learning pathways, provide full online capability, and add collaborative learning modes where students could work together to solve a problem.In conclusion, Propolingo is a language learning application modelled after Duolingo designed to address the challenges faced by undergraduate students in South Africa. By incorporating gamification techniques, interactive exercises, and a user-focused design, Propolingo not only boosts student engagement but provides access to propositional logic in an accessible format which can supplement the traditional computer science curriculum in South Africa, which does not cover this important logic essential concepts. Its adaptability allows learners to advance at their own pace, offering a personalized learning experience that aligns with individual needs and curriculum goals.

References

1. Ben-Ari, M.: Mathematical logic for computer science. Prentice-Hall, Inc., USA (2013)
2. Maharaj, A.: Propositional logic and ai education in south africa. Journal of Educational Research in Mathematics and Science **6**(2), 102–115 (2014)
3. Russell, S., Norvig, P.: Artificial Intelligence: A Modern Approach. Prentice Hall, 3 edn. (2010)

Appendix

GitHub Repository

The source code for the *Propolingo* application can be found at the following GitHub repository:

- https://github.com/Baleni-BG/Propolingo

"Can an Intelligent Robot Inherit Intellectual Property?"

Technological Development vs EU AI Act. Human-centred World Vision.

THE FUTURE IS NOT TOMORROW - THE FUTURE IS TODAY

dr. Csilla Mónika Rácz

[0009-0004-3524-7887]

University of Szeged- Hungary, Faculty of Law and Political Science Doctoral School

(csillamonikaracz@gmail.com)

Supervised by: Prof. Dr. Csongor István Nagy Prof. Dr. Csongor István Nagy Ph.D., LL.M., S.J.D.

and Dr. Gábor Hajdu Senior Lecturer, Ph.D. (doktori.iskola@juris.u-szeged.hu)

Abstract. Can an intelligent robot inherit intellectual property? The relevance of this question is that on 1 August 2024 came into force the world's first regulation of Artificial Intelligence, the European Union's AI Act (hereinafter the AI Regulation). In this research I compare the current state of protection of human intellectual property rights against AI under the patent laws of the European Union, Hungary, Switzerland, South Africa and Saudi Arabia. Thought-provoking questions also: -Have the legislators built sufficient guarantees into the AI Regulation to protect intellectual property as a technological development for the benefit of Man against artificial intelligence? -Is the inheritance, succession and succession of rights adequately ensured under the regulation if AI is involved in the creative process? -What happens if the inventor's invention is cannot obtain patent protection, because it is not the intellectual product of Man, but has resulted from the autonomous activity of the intervening artificial intelligence? Few years ago, these would have been absurd, incomprehensible questions, but today, when intelligent robots can be granted human constitutional rights, every nation must realize:

THE FUTURE IS NOT TOMORROW - THE FUTURE IS TODAY

Keywords: artificial intelligence, inheritance law, EU AI Act, investment law, technology- transfer

1 Introduction

On 1 August 2024, the World's first legislation on artificial intelligence, the European Union's AI Act, came into force. [1] Theoretically, this could be the world's shortest treatise, since the first sentence of the Regulation already provides the answer to the question:

"The fundamental purpose of the AI Regulation is to promote the spread of human-centred, trustworthy AI."

In doing so, the legislator has made it clear that the millennia-old European values of human-centredness are still incompatible with the vision of "Robo Sapiens"[2].

However, some more thought-provoking questions may arise which are worth briefly exploring:

-why was EU-level AI regulation timely,
-who is entitled to the property or property rights resulting from the exploitation and use of the invention if a person cannot prove his/her inventive capacity or the machine is designated as the inventor. Also very important to know: is it safe to invest in AI development? Can AI inherit intellectual property, can it be heir, legator?
A few years ago, these would have been absurd, unintelligible, or fanciful fantasy questions, but today, when intelligent robots can be given human constitutional rights, every nation must realize that 'THE FUTURE IS NOT TOMORROW - THE FUTURE IS TODAY.'

2 Research areas and Method

This paper is an extract from a larger, more detailed study in which I compared the current state of protection of human intellectual property rights against AI under EU, Hungarian, Swiss, South African and Saudi patent law.
I examined why was EU-level AI regulation timely? Therefore, I compared the following areas of law: Intellectual property law – patent law, inheritance law, Technology transfer as a specific area of intellectual property law and the Investment Law FROM THE PERSPECTIVE OF AI. It is worth pondering that nowadays an intelligent robot be given constitutional human rights ('Sofia' in Saudi Arabia) [3] and also AI has already been certified as inventor (The Dabus case) [4] whether the legislator should be prepared for the possibility that AI could be the heir to intellectual property.

Can AI be heir if outlives man, or can be legator if dies out of the ranks of robots?

To answer this question, I compared the intellectual property law of the European Union's Artificial Intelligence Regulation with the patent laws and intellectual property law of Hungary, Switzerland, the Republic of South Africa and Saudi Arabia. [5-31]

1. Why was EU-level AI regulation timely?
2. Intellectual property law and inheritance law from the perspective of AI.
3. Technology transfer as a specific area of intellectual property law and the AI.
4. The relationship between intellectual property law and succession law– Comparison of Hungarian and EU legislation
5. Intellectual Creations as Digital Innovation - Investment Law - and AI
6. European Union AI Regulation - The intellectual property regulation and rights The relationship between AI and Human in the AI Regulation
7. Outside the EU, but in the heart of Europe -Swiss AI snapshot
8. International trends, legal cases from the World of AI.
9. The Dabus case and rejection of AI inheritance in intellectual property.
10. The Republic of South Africa and the AI Intellectual Property Inheritance Vision
11. Comparison of Republic of South African and Hungarian patent law
12. The new AI regulatory ambitions of the Republic of South Africa
13. Saudi Arabia's intellectual property law in relation to the AI

3 Conclusion

Under current European Union and Hungarian law, AI cannot acquire intellectual property rights, nor can be an heir or legatee, as nor under South African, Saudi and Swiss law. In my opinion, the AI Regulation has built in basic guarantees for the protection of intellectual property as a technological development for the benefit of man as opposed to artificial intelligence, but the specific regulation of this is the task and responsibility of the Member States. It is also the responsibility of the Member States to ensure the safe inheritance and transmission of intellectual property, the succession of rights - in accordance with the intellectual property Regulation - for the benefit of Man, if AI is involved in the creative process. If the inventor's invention is not patented, because it cannot be patented on the grounds that it is not his/her intellectual product, but the result of the autonomous activity of the contributing AI, - this remains an open question, as the EU AI Act. does not specifically mention it. However, Article 14(1) provides that "Human supervision is necessary, high-risk AI systems shall be designed and developed, including by appropriate human-machine interface tools, in such a way that they can be effectively supervised by natural persons during the period of their use." This suggests that it is worth investing in AI both in the EU and in Hungary, and other countries around the world are also striving to provide the right environment for businesses to do so.

4 References

1. REGULATION (EU) 2024/1689 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 13 June 2024 laying down harmonised rules for artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139, (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (the Artificial Intelligence Regulation) https://commission.europa.eu/news/ai-act-enters-force-2024-08-01_en
2. Tamás Klein: Robot law or human law? In Bernát Török and Zsolt Ződi (eds.): The regulatory challenges of artificial intelligence. Studies on artificial intelligence and its frontiers. Ludovika University Press. Budapest, 2021.
3. Thaler v Comptroller-General of Patents, Designs and Trademarks [2023] UKSC 49. Supreme Court UK. https://www.supremecourt.uk/cases/uksc-2021-0201.html
4. Kinga Kálmán: The case of the world's first robot citizen - or the new challenge of legislation. Arsboni 2018.11.05. https://arsboni.hu/a-vilag-elso-robot-allampolgaranak-esete-avagy-a-jogalkotas-ujabb-kihivasa/
5. Bernard Marr: Understanding the 4 types of artificial intelligence. https://bernardmarr.com/understanding-the-4-types-of-artificial-intelligence/
6. Levente Tattay-Pogácsás Anett-Ujhelyi Dávid: Intellectual Property Rights. Szent István Társulat The Book Publishing House of the Holy See of the Apostolic See Budapest, 2021.
7. European Commission: Shaping Europe's digital future -AI Act. https://digitalstrategy.ec.europa.eu/en/policies/regulatory-framework-ai
8. Eden Winlow: The end of the road for DABUS and Dr Thaler in the UK Supreme Court Kluwer patent blog (Bristows)/2024 16 January 2024. https://patentblog.kluweriplaw.com/2024/01/16/the-end-of-the-road-for-dabus-and-dr-thaler-at-the-uk- supreme-court/
9. Thaler v Comptroller-General of Patents, Designs and Trademarks [2023] UKSC 49. The Supreme Court UK <https://www.supremecourt.uk/cases/uksc-2021-0201.html >
10. John McCarthy (1927- 2011) American mathematician and computer scientist, pioneer of artificial intelligence (AI) <https://www.britannica.com/biography/John-McCarthy>
11. István Molnár: Intellectual Property Management and Technology Transfer. Innov AID Innovation and Gazdasági Tanácsadó Kft. Kecskemét 2008. Lakos Printing House. ISBN 978-963-06-6396-0.
12. Technology transfer <https://hu.economy-pedia.com/11039552-technology-transfer >
13. András Koltay (foreword) In Bernát Török and Zsolt Ződi (eds.): A mesterséges intelligencia szabályozási kihívásai. Research on the regulatory challenges of artificial intelligence. Ludovika University Press.

14. Kinga Hazai: Once upon a time there was a creator. Once upon a time there was a man. Hungarian Journal Budapest, 2023.
15. 100 Hungarian Inventions and Developments 2021. Another 100 Hungarian innovations have been added to the TOP 100 Media series. https://energiaoldal.hu/megjelent-a-100-magyar-talalmany-esfejlesztes-2021- cimu-kiadvany/
16. Hungarian researchers help serious patients get a new kidney sooner ORIGO 2024.08.05.
17. https://www.origo.hu/tudomany/2024/08/meg-iden-elindul-a-magyar-vesecsereprogram
18. NEWS - Artificial Intelligence News. Artificial intelligence could be a breakthrough for the Hungarian economy Source: MTI 2023.11.30. 13:14 < https://hirado.hu/extra/tudomany-high-tech/cikk/2023/11/30/a-mesterseges-intelligencia-a-magyarhost-guest-guest-point-possible >
19. The government turns to artificial intelligence Author. editor 2024. június 3. < https://hirlevel.egov.hu/2024/06/03/rafordul-a-kormany-a-mesterseges-intelligenciara/>
20. ITIDA Organizes Egyptian-Hungarian ICT Business Forum to Forge Partnerships and Joint Investments. KAIRO 2 June 2024.
21. < https://itida.gov.eg/English/PressReleases/Pages/ITIDA-Organizes-Egyptian-Hungarian-ICT-Business-Forum.aspx >
22. Ben Carlson: Investing with common sense. Why is it (also) better to be simple than complex in financial markets? (trans. Szabolcs Lénárt) A4C Books, Budapest, 2017.
23. Bernard Marr: Understanding the 4 types of artificial intelligence. < https://bernardmarr.com/understanding-the-4-types-of-artificial-intelligence/ >
24. Andre Kostolany: Bilanz der Zukunft (translated by Dóra Svéda Perfekt Kiadó Rt. Budapest, 2016. Original edition: Andre Kostolany's Bilanz der Zukunft. This morning. Published in 1995 by Econ Verlag GmbH Düsseldorf.
25. Dr. Dániel Necz: The impact of master intelligence on the performer's right. <https://www.sztnh.gov.hu/sites/default/files/files/kiadv/szkv/szemle-2018-06/02.pdf >
26. Peter Stanyer: Investing. From safe solutions to high risk. HVG Publishing Ltd. Budapest, 2008
27. Attila Marján: European Finance. Banks, stock exchanges, the single currency and global competition. Attila Marján refers to the Lámfalussy report (Fial report of the Committee of Wise Men on the reputation of European securities markets, 15 February 2001 Brusseles).
28. Tattay Levente-Pogácsás Anett-Ujhelyi Dávid. Szent István Társulat The Book Publishing House of the Holy See of the Apostolic See Budapest, 2021. p. 32.
29. Tattay Levente-Pogácsás Anett-Ujhelyi Dávid (2021). p. 33.
30. The World Intellectual Property Organization (WIPO) is the United Nations organisation that serves the world's innovators and creators, ensuring that their ideas can safely reach the marketplace and improve lives https://www.wipo.int/about-wipo/en/
31. Source: WIPO Statistics Database, March 2024. https://www.wipo.int/en/ipfactsandfigures/patents

Spatio-Temporal Graph Neural Networks for Human Activity Recognition

Suvana Rohanlal[1,2] and Deshendran Moodley[1,2]

[1] University of Cape Town, Cape Town, South Africa
[2] Centre for Artificial Intelligence Research (CAIR), South Africa
RHNSUV001@myuct.ac.za | deshen@cs.uct.ac.za

Abstract. The proliferation of sensor based monitoring devices has furthered the development of Human Activity Recognition models. In this research, we perform a comparison between three deep learning models: Bi-Directional Long Short-Term Memory(Bi-LSTM), Temporal Convolution Networks(TCN) and Spatio-Temporal Graph Convolutional Networks(STGCN), for the classification and prediction of activities of daily living using the CASAS dataset.

Keywords: Human Activity Recognition · Digital Twins · Deep Learning · STGNN

1 Introduction

Human Activity Recognition (HAR) algorithms are used to accurately classify activities of daily living. Several recent studies have explored the use of deep learning models such as Long Short-Term Memory (LSTM) [7], Temporal Convolution Networks (TCN) [11][12], Graph Convolution Networks (GCN) [1] for the classification of human activities. Although Spatio-Temporal Graph Neural Networks(STGNNs) have been used for skeleton-based activity recognition, they have not yet been explored in the context of smart home-based recognition [8][9].

This research introduces a novel approach to HAR, based on smart homes and STGNNs for the classification and prediction of activities of daily living. This approach is a Spatio-Temporal fusion neural architecture for STGNNs, called Spatio-Temporal Graph Convolutional Networks (STGCN) [8]. This implementation is evaluated against two baseline models; Bi-LSTM [7] and TCN [11], using a dataset from the Center of Advanced Studies in Adaptive System (CASAS) collection [4].

2 The HAR problem

Human Activity Recognition can be formulated as a classification problem, where, data instances are a set of temporal data sequences, collected over a time period using sensors.

Given a fixed sliding window over all n amount of sensors; $S_i = \{S_{i,0}, ..., S_{i,n}\}$, and a set of activity labels; $A = \{a_0, ..., a_k\}$. The aim is to find a function to map $f : S_i -> A$ for all possible values of S_i, where $f(S_i)$ represents the activity performed during the window [6][10].

Two particularly challenging aspects of HAR classification include the the identification of new activities and the prediction of future activities [2].

3 Approaches

Spatio-Temporal Graph Convolutional Network: The framework presented in this research consists of multiple Spatio-Temporal Convolutional blocks, each containing one GCN learning network sandwiched between two TCN learning networks [5]. The sensors are represented as nodes, thus forming the structure of a graph. This allows meaningful spatial patterns and features to be extracted when the graph data is passed through the graph convolution [13]. The temporal features are extracted through convolutions resulting in a parallel design. This means that all information of a given time window is input at the same time due to the convolutional structures [5]. The learning model is then connected to a Softmax layer for classification [8].

Baseline models: This research employs two distinct baseline implementations for comparative analysis; the Bi-LSTM [7] represents the state-of-the-art performance on the CASAS dataset, while the TCN [11] reflects the state-of-the-art performance in the field of smart home based HAR. These models will be evaluated using accuracy, precision, recall and F-score.

4 Conclusion

It is expected that the combination of the TCN and GCN models with a Softmax output layer will provide comparable results to the current state-of-the-art implementations.

As an extension, a digital twin will be designed to augment the CASAS dataset with the aid of the AI system. Current digital twin implementations demonstrate agents that perform activities of daily living to replicate datasets [3], whereas our implementation will deploy self-learning agents to enhance datasets. This system will be designed to continuously learn and adapt according to the initial data provided. The AI system will showcase the agent performing activities of daily living within the confines of a virtual model of a physical apartment.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Ahmad, T., Jin, L., Zhang, X., Lai, S., Tang, G., Lin, L.: Graph convolutional neural network for human action recognition: A comprehensive survey. IEEE Transactions on Artificial Intelligence **2**(2), 128–145 (2021). https://doi.org/10.1109/TAI.2021.3076974
2. Bi, H., Perello-Nieto, M., Santos-Rodriguez, R., Flach, P.: Human activity recognition based on dynamic active learning. IEEE Journal of Biomedical and Health Informatics **25**(4), 922–934 (2020)
3. Bouchabou, D., Grosset, J., Nguyen, S.M., Lohr, C., Puig, X.: A smart home digital twin to support the recognition of activities of daily living. Sensors **23**(17), 7586 (2023)
4. Cook, D.J., Crandall, A.S., Thomas, B.L., Krishnan, N.C.: Casas: A smart home in a box. Computer **46**(7), 62–69 (2012)
5. Jin, G., Liang, Y., Fang, Y., Shao, Z., Huang, J., Zhang, J., Zheng, Y.: Spatio-temporal graph neural networks for predictive learning in urban computing: A survey. IEEE Transactions on Knowledge and Data Engineering pp. 1–20 (2023). https://doi.org/10.1109/TKDE.2023.3333824
6. Lara, O.D., Labrador, M.A.: A survey on human activity recognition using wearable sensors. IEEE Communications Surveys Tutorials **15**(3), 1192–1209 (2013). https://doi.org/10.1109/SURV.2012.110112.00192
7. Liciotti, D., Bernardini, M., Romeo, L., Frontoni, E.: A sequential deep learning application for recognising human activities in smart homes. Neurocomputing **396**, 501–513 (2020)
8. Lovanshi, M., Tiwari, V.: Human skeleton pose and spatio-temporal feature-based activity recognition using st-gcn. Multimedia Tools and Applications **83**(5), 12705–12730 (2024)
9. Min, S., Gao, Z., Peng, J., Wang, L., Qin, K., Fang, B.: Stgsn—a spatial–temporal graph neural network framework for time-evolving social networks. Knowledge-Based Systems **214**, 106746 (2021)
10. Minor, B., Cook, D.J.: Forecasting occurrences of activities. Pervasive and Mobile Computing **38**, 77–91 (2017). https://doi.org/https://doi.org/10.1016/j.pmcj.2016.09.010
11. Nair, N., Thomas, C., Jayagopi, D.B.: Human activity recognition using temporal convolutional network. In: Proceedings of the 5th international Workshop on Sensor-based Activity Recognition and Interaction. pp. 1–8 (2018)
12. Wei, X., Wang, Z.: Tcn-attention-har: human activity recognition based on attention mechanism time convolutional network. Scientific Reports **14**(1), 7414 (2024)
13. Yu, B., Yin, H., Zhu, Z.: Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. arXiv preprint arXiv:1709.04875 (2017)

Exploring Implementations of KLM-Style Defeasible Explanation Algorithms

Chipo Hamayobe[0009-0008-0987-814X] * ** ***

University of Cape Town
chipo@cs.uct.ac.za

1 Introduction

Defeasible reasoning closely mirrors human reasoning processes as it allows us to represent information that may include exceptions [6]. There are various approaches to this type of reasoning, but we focus only on inferences within the *KLM approach*, proposed by Kraus, Lehmann, and Magidor (KLM) [7].

Beyond the ability to infer new information, it is equally important to offer explanations as to why certain inferences hold. In this study, we focus on *justifications*, a well-established and conceptually straightforward form of explanation in defeasible reasoning [5].

The aim of our project is to explore, develop, implement, and evaluate explanation algorithms for KLM-style inference operators, namely *Rational Closure*, *Lexicographic Closure* and *Relevant Closure*. We aim to conduct a comparative analysis to reveal the trade-offs between these inference operators in terms of computational complexity, entailment strength, and justification quality. We hope the experimental evaluations will demonstrate and reveal the efficacy of the algorithms in real-world applications, highlighting the algorithms's balance between computational efficiency and entailment strength. This research contributes to the development of scalable and robust defeasible reasoning systems.

2 Defeasible Entailment

The KLM framework extends propositional logic by introducing *defeasible implication* $\mid\sim$, the defeasible counterpart to the classical implication $\rightarrow$. These are expressions of the form $\alpha \mid\sim \beta$, read as *'α typically implies β'*, where $\alpha, \beta \in \mathcal{L}$ (set of all well-formed statements). A defeasible knowledge base $\mathcal{K}$ consists of a finite set of such defeasible implications, and defeasible entailment $\mid\approx$ is a binary relation between defeasible knowledge bases and defeasible implications, such that $\mathcal{K} \mid\approx \alpha \mid\sim \beta$ means *'$\mathcal{K}$ defeasibly entails that α typically implies β'*.

Lehmann and Magidor [9] introduced a set of postulates that characterise *rational* defeasible entailment, with each postulate reflecting an intuitive property

* Supervised by Professor Tommie Meyer (tmeyer@cs.uct.ac.za)
** Co-Supervised by Dr Jonas Haldimann (jonas@haldimann.de)
*** Supported by the Artificial Intelligence Research Unit (www.airu.org.za)

expected of a coherent defeasible entailment relation, thereby justifying the term *rational*. Beyond axiomatic definitions, although not the focus of this discussion, these relations also have model-theoretic semantics-based expressions. We focus precisely only on reasoning algorithms with tractable computational complexity [4].

2.1 Rational Closure (RC)

Proposed by Lehmann and Magidor [9], RC is a rational formalisation of defeasible entailment. Casini et al. [2] provides algorithms through a two-stage distinct process, `BaseRank` and `RationalClosure`. The former ranks statements in $\mathcal{K}$ in order of typicality; while the latter, when confronted with inconsistencies based on the antecedent of the query, removes the most typical rank.

2.2 Lexicographic Closure (LC)

LC constitutes an alternative rational framework for defeasible entailment, initially proposed by Lehmann [8], which exhibits greater permissiveness compared to RC and can also be articulated both semantically and algorithmically. LC functions as a refinement of Rational Closure by eliminating individual statements rather than entire hierarchical levels when inconsistencies emerge during the reasoning process.

2.3 Relevant Closure (RelC)

Proposed by Casini et al. [1], RelC is a modified inference operator that builds upon RC. By restricting retractions to statements in lower specificity ranks that directly contradict more specific statements in higher ranks with respect to the query's antecedent, RelC seeks to enhance the precision of the reasoning process. Although RelC is not rational and hence deviates from rational defeasible entailment by not adhering to all its axioms, it offers a more targeted approach to defeasible reasoning [4].

3 Defeasible Explanations

Explanations give reasons for why an entailment holds from a knowledge base. This project focuses on a specific type of explanation called *justifications*, which are minimal subsets of formulas needed for an entailment to hold [5].

Suppose we have $\mathcal{K} = \{b \mid\sim w, b \mid\sim f, t \rightarrow b\}$ and a query $t \mid\sim f$? The only statements needed to answer the query are $\mathcal{J} = \{b \mid\sim f, t \rightarrow b\}$. The statement $b \mid\sim w$ is irrelevant in this case because the entailment still holds without it but not without any $\alpha \in \mathcal{J}$. $\mathcal{J}$ is the justification for $\mathcal{K} \approx t \mid\sim f$.

Chama [3] proposed a RC justification algorithm for Defeasible *Description Logics* that was adapted to rules based on *Propositional Logics* by Wang [10]. We plan to use this basis and also adopt the algorithms proposed by Everett et al. [4] as a foundation for the approach to justification algorithm development for LC and RelC inferences.

References

1. Casini, G., Meyer, T., Moodley, K., Nortjé, R.: Relevant closure: A new form of defeasible reasoning for description logics. In: Logics in Artificial Intelligence: 14th European Conference, JELIA 2014, Funchal, Madeira, Portugal, September 24-26, 2014. Proceedings 14. pp. 92–106. Springer (2014)
2. Casini, G., Meyer, T., Varzinczak, I.: Taking defeasible entailment beyond rational closure. In: Logics in Artificial Intelligence, pp. 182–197. Lecture Notes in Computer Science, Springer International Publishing, Cham (2019)
3. Chama, V.: Explanation for defeasible entailment. Master's thesis, Faculty of Science, University of Cape Town, Rondebosch, Cape Town, 7700 (2020)
4. Everett, L., Morris, E., Meyer, T.: Explanation for KLM-Style Defeasible Reasoning. In: Southern African Conference for Artificial Intelligence Research. pp. 192–207. Springer, South Africa (2021)
5. Horridge, M., Parsia, B., Sattler, U.: Justification masking in owl. In: 23rd International Workshop on Description Logics DL2010. p. 32 (2010)
6. Kaliski, A.: An Overview of KLM-Style Defeasible Entailment. Master's thesis, Faculty of Science, University of Cape Town, Rondebosch, Cape Town, 7700 (2020)
7. Kraus, S., Lehmann, D., Magidor, M.: Nonmonotonic reasoning, preferential models and cumulative logics. Artificial Intelligence **44**(1), 167–207 (1990). https://doi.org/10.1016/0004-3702(90)90101-5
8. Lehmann, D.: Another perspective on default reasoning. Annals of Mathematics and Artificial Intelligence **15** (11 1999). https://doi.org/10.1007/BF01535841
9. Lehmann, D., Magidor, M.: What does a conditional knowledge base entail? Artificial Intelligence **55**(1), 1–60 (1992). https://doi.org/10.1016/0004-3702(92)90041-U
10. Wang, S.: Defeasible Justification for the KLM Framework. Master's thesis, Faculty of Science, University of Cape Town, Rondebosch, Cape Town, 7700 (2022)

Data-driven LSTM power grid inertia estimations

Ethan Codron[1], Thomas Drageset[2] Leonardo Rydin Gorjão[2[0000-0001-5513-0580]], P. Armand Bester[1], and Jacques Maritz[1[0000-0003-1556-8523]]

[1] Department of Engineering Sciences, University of the Free State, Bloemfontein 9301, South Africa
[2] Faculty of Science and Technology, Nordic University of Life Sciences, 1432 Ås, Norway

Abstract. Estimating the inertia in a power system at any given time is a complex process, especially for large interconnected grids that rely on a diverse mix of generation and trading across borders. Estimations of the total inertia in a power system permit system operators to ensure system stability, e.g. by allocating extra resources in low-inertia states. Determining the exact, ground truth, inertia level is impossible, especially for power systems where the individual inertia types are not easily accessible. Most works focus on synchronous generator models and numerical simulations. In this work, we explore the ability of a recurrent neural network, namely Long Short-Term Memory (LSTM), to capture inertia from extant data. We illustrate the influence of the 'ground truth' on such an inertia estimator for the Nordic grid and outline the limitations when attempting to apply such a model for inertia estimations associated with the South African power grid.

Keywords: Stored kinetic energy · inertia estimation · purely data-driven model · LSTM

1 Introduction

Power systems are one of the most complex systems on earth, ultimately being a lifeline of modern society. Ensuring the stability of these systems is crucial and requires accurate state estimations of the dynamics governing the systems [8, 5]. Grid operators are specifically concerned with the total existing kinetic energy stored in large synchronous generators that drive a fine balance between generation and demand [5, 7]. Mixing generation types, i.e., inertial and non-inertial, complicates the assessment and estimation of available kinetic energy stored. From a transmission system operator or a country or region, active importing and exporting power further hinders these estimations as it is impossible to distinguish the power source types. and highly transient locally dependent generation states, generate inaccuracies in estimating the total inertia at a given time in the power system [5]. One estimation technique involves assigning generic inertia constants to each generation type in the power system [9]

$$E_{k,sys} = \sum_{m=1}^{N} S_m H_m = \sum_{m=1}^{N} \frac{P_m}{\text{p.f.}} H_m \tag{1}$$

where H_m is a constant for each production type, m, P_m is the total active power, S_m is the total apparent power and p.f. is the power factor. On short time scales, a typical power system (and its generation subsystems) will keep the power factor constant and vary the total active power based on generation needs. The latter makes for an aggregated approach to estimating the total inertia in the system and depends heavily on active power data and the mix of generation types. Since Eq. (1) is a linear process, the applicability of machine learning techniques becomes appealing given the existent data of power generation by type [3, 4]. Particularly, the application of recurrent neural networks, particularly Long Short-Term Memory (LSTM) models, in time series forecasts [6].

The rest of the paper is structured as follows: Section 2 provides a framework to utilise the LSTM-based regression model for the estimation of power system inertia based on some different ground truths (actual measurements *vs* estimations), Section 3 outlines the performance of the LSTM based model for different sets of ground truths for Norway power system, Section 4 crystallises the implications of an uncertain ground truth on the future process of estimating inertia in SA power system, Section 5 concludes the study with some recommendations.

2 Methods and data

The power generation data of South Africa (SA) were obtained from the ESKOM data portal [4]. Eq. (1). The power generation data for the Nordic power systems was obtained from the ENTSO-E data portal [3] and Energinet for the kinetic energy measurements [2]. Eq. (1) and the inertia constants (see Tab. 1) were used to estimate the total inertia in the SA power systems for the year 2023 and were compared to the Nordic (NW) inertia estimates for the same year (cf. Fig. 1). Ultimately, comparing two sets of ground truths that were generated using generic inertia constants. The generation type decomposition for the major inertia contributors is highlighted in Fig. 2.

Table 1. Generic constants used in inertia estimations applicable to both power systems [9].

Generation Type	$H[s]$	pf
Nuclear	6.3	0.9
Other Thermal	4.0	0.9
Hydro-conventional	3.0	0.9
Hydro-small-scale	1.0	0.9
Wind and PV	0.0	0.9

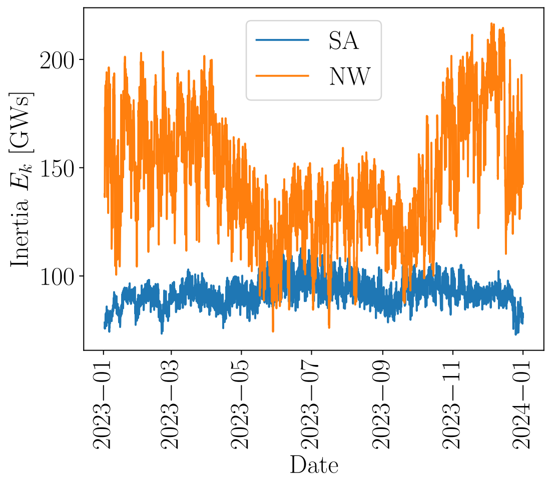

Fig. 1. Comparison of total inertia (estimated via generic inertia constants) for the SA and NW power systems for 2023. Seasonal variations can be observed, with a pronounced inertia during the Winters, which are opposite given each location is on a different hemisphere.

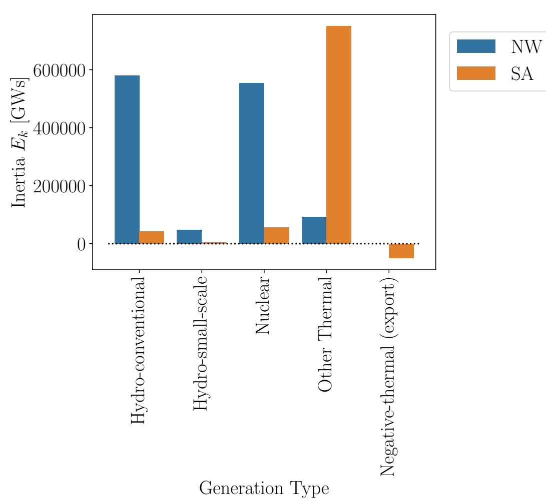

Fig. 2. Cumulative kinetic energy decomposition for the SA and the NW power systems based on generation type and exports for 2023. The systems operate very distinct generation types.

The above generically calculated inertia forms the ground truth for both the SA and NW power systems and becomes the insertion point to investigate the ability of LSTM regression models to estimate the inertia of a power system and the use of inertia constants. The latter also entertains the possibility of training on one power system and predicting (at least global signatures) for another power system. When observing Eq. (1), it becomes clear that the chosen training model needs to conform to a sequential model.

In this paper, the estimation (or regression) of the inertia of the power system is based on an LSTM model. Given the rather limited number of generation types (20), the LSTM model considered is comprised of a single LSTM layer.

Table 2. Architecture of the machine learning model used to estimate inertia in both power systems

Layer	Number of Nodes	Trainable parameters
Input	20	0
LSTM	40	9760
Dense	1	41

The layered network outlined in Tab- 2 above, was implemented in the Keras package from within Tensorflow [1]. The model takes the shape of a sequential model joining a dense and LSTM layer. The input layer is mapped to the dimensions of the training data and the transfer occurs via a 40-node LSTM inter-layer and converges to a single value via a dense layer.

Prepossessing included feature selection-based generation decomposition and a total of 20 features were used pertaining to the generation mix of the Nordic power system. For training on the Nordic inertia data, the LSTM was granted knowledge of the daily cycle that reflects the power generation data. The training data was that of hourly inertia values that were retrieved from Ref. [2] between 2020 and 2023.

3 Results and discussions

Fig. 3 showcases the LSTM forecast for a 1-week selection of inertia samples, totalling 168 hours, from the Nordic training set (2020–2023), ultimately containing a more accurate ground truth that was generated from actual inertia measurements. Fig. 4 similarly shows a forecast taken from the 2023 Nordic data set generated using generic inertia constants, which constitute a less accurate ground truth, given the simplification to 5 generation types. In this paper, the benchmark used to evaluate the random sample estimations against the targets is the R2 score.

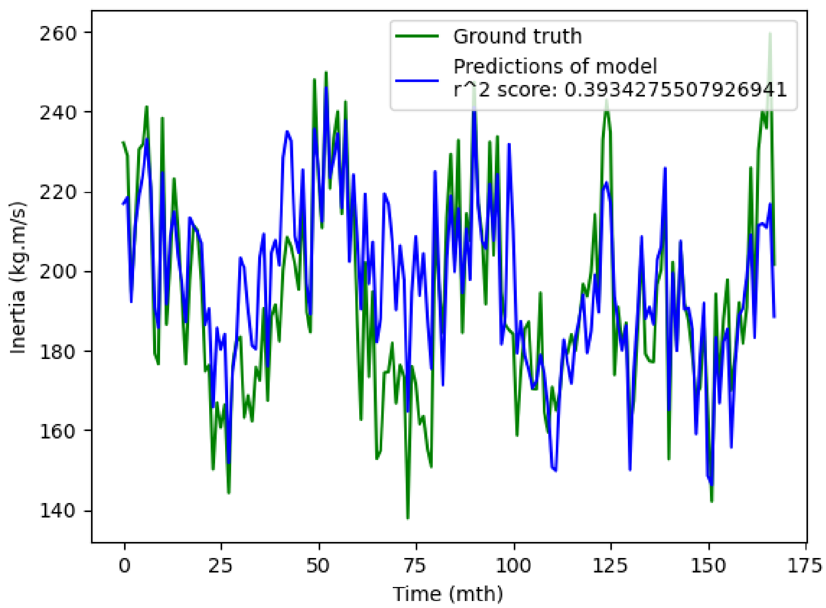

Fig. 3. Regression attempt for random inertia samples, totalling 168 hours, from the Nordic training set (2020–2023).

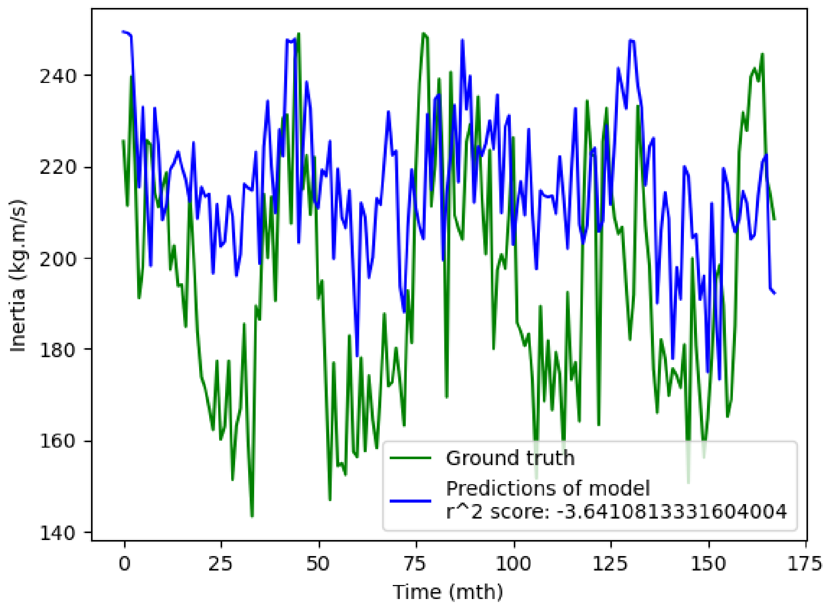

Fig. 4. Regression attempt for random inertia samples inertia, totalling 168 hours, taken from the 2023 Nordic data set that was generated using generic inertia constants (less accurate ground truth).

4 Recommendations and Conclusions

The main difference between inertia regression for the random samples as observed in Fig. 3 and Fig. 4, specifically pertaining to the NW grid, is the presence of a more accurate ground truth, that is, inertia retrieval via measurement versus inertia estimated using generic inertia constants and active power generated by the power system. This agrees with what is expected in large power systems. Only the transmission system operators (TSOs) have access to the actual inertia of each synchronous generator. Nevertheless, access to boiled-down generation-by-type data, as presented in this work, can serve for inertia estimation of a large class of power systems, for as long as data is available. Machine-learning applications, like the one carried out in this work, can learn from the more accurate inertia 'ground truth' provided by the TSOs and connect these to less detailed data. This serves as a bedrock for estimations wherein TSOs do no have or provide inertia estimations (e.g. in SA).

References

1. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D.G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu, Y., Zheng, X.: Tensorflow: A system for large-scale machine learning. 2016, Last Accessed: 2024-04-16 https://arxiv.org/abs/1605.08695
2. Energinet: Inertia, nordic synchronous area. Last Accessed: 2024-04-23 https://www.energidataservice.dk/tso-electricity/InertiaNordicSyncharea
3. ENTSO-E: Entso-e – actual generation per production type, https://transparency.entsoe.eu/generation/r2/actualGenerationPerProductionType/show
4. ESKOM: Eskom – data request, https://www.eskom.co.za/dataportal/data-request-form/
5. Heylen, E., Teng, F., Strbac, G.: Challenges and opportunities of inertia estimation and forecasting in low-inertia power systems. Renewable and Sustainable Energy Reviews **147**, 111176 (2021). https://doi.org/10.1016/j.rser.2021.111176
6. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Computation **9**(8), 1735–1780 (1997). https://doi.org/10.1162/neco.1997.9.8.1735
7. Prabhakar, K., Jain, S.K., Padhy, P.K.: Inertia estimation in modern power system: A comprehensive review (10 2022). https://doi.org/10.1016/j.epsr.2022.108222
8. Ulbig, A., Borsche, T.S., Andersson, G.: Impact of low rotational inertia on power system stability and operation. IFAC Proceedings Volumes **47**, 7290–7297 (2014). https://doi.org/10.3182/20140824-6-ZA-1003.02615
9. Ørum, E., Kuivaniemi, M., Laasonen, M., Bruseth, A.I., Jansson, E.A., Danell, A., Elkington, K., Modig, N.: Future System Inertia. Last Accessed: 2024-04-16 (2015), https://eepublicdownloads.entsoe.eu/clean-documents/Publications/SOC/Nordic/Nordic_report_Future_System_Inertia.pdf

Utilizing Hybrid LSTM-CNN Deep Learning Models for Heart Disease Prediction

Abstract

Heart diseases are among the world's major causes of death, and their early diagnoses are among the major reasons for reduced mortality. In recent times, Machine Learning has shown some prospects in aiding health professionals in heart disease diagnoses at an early stage. This research proposes a new hybrid deep learning algorithm by fusing Long Short-Term Memory networks and Convolutional Neural Networks for cardiovascular disease diagnosis. This hybrid model contributes to an improvement in the accuracy of prediction through the strength of the long short-term memory (LSTM), which captures temporal patterns, and that of CNN, which is excellent for extracting spatial features from clinical data. The demographic and clinical data of age, blood pressure, cholesterol levels, and ECG readings were collected from a high-risk population in the Western Cape, South Africa.

It involves cleaning up the data in the pre-processing stage where missing values are treated with the help of normalization techniques so that the data keeps on the same scale. The obtained dataset was further divided into training and testing sets in the ratio 70:30. The proposed hybrid model has used the CNN layer to process raw data for its critical feature extraction, while the LSTM layer is used to model temporal dependencies present within the sequential data. The result of the output was a binary classification showing whether heart disease was present or not. It was compared to several machine learning models' performance, including support vector machines, k-nearest neighbor, and Naive Bayes classifiers.

For the proposed hybrid model, the accuracy is 89%, sensitivity is 81%, and specificity is 93%. These results confirm that the hybrid LSTM-CNN model is best for predicting heart disease, with significant improvements in the accuracy and generalization capability compared to the models previously proposed. Furthermore, the confirmation of its ability to show clear differentiation between heart disease-positive and negative cases was done with an AUC score of 0.86.

It serves to emphasize that deep learning models, more so hybrid architectures, may offer great promise in the early detection of heart disease, and indeed the latter proves to be an asset in clinical decision-making. Further work may be conducted to include real-time medical datasets and the applicability of the proposed model in other conditions affecting the cardiovascular system.

Furthermore, the performance of the hybrid LSTM-CNN model has shown that deep learning outperforms traditional machine learning in complex medical data analysis. Temporal dynamics in the clinical readings, such as a change in blood pressure or cholesterol levels over time, are well represented by the LSTM component, while spatial patterns in features like ECG signals are well recognized with CNN layers. This dual approach gives a more holistic analysis of patient data for better prediction.

It also points to the most interesting aspect of being cautious with the preprocessing data, such as how to handle missing values and normalize data for the robustness of a predictive model. While treating missing data points and normalizing the clinical indicators, the research has ensured that the model effectively operates on real-world datasets that usually contain irregularities. Besides this, the performance of the model has been validated on data it did not see when it was trained on a 70-30 split of the dataset, and because of this, it makes this model an appropriate candidate for a clinical setting.

This hybrid model has given the best results in all three important metrics-accuracy, sensitivity, and specificity-compared to other models such as support vector machines, k-nearest neighbors, and Naive Bayes. It tends to show that CNN-LSTM works very well in heart disease prediction because it considers both spatial and temporal features of data, which are critical in understanding cardiovascular conditions.

The proposed hybrid LSTM-CNN model presents an effective method for the prediction of heart disease by improving diagnostic accuracy and enhancing its reliability. Capable of handling diverse patient data, it offers clinically relevant and interpretable results, making this architecture one of the most promising for healthcare professionals. Further research is needed regarding the performance of the model on real-time data and large datasets, and its application to other diseases with similar characteristics should also be explored. Refining similar deep learning models could become a game-changer in cardiovascular health, one day providing timely and appropriate diagnosis to improve patient outcomes worldwide.

Keywords:

Heart Disease Prediction, Hybrid LSTM-CNN, Deep Learning, Machine Learning, Clinical Data