Journal of Telecommunications and the Digital Economy

Volume 12, Number 3
September 2024

Special Issue:

Perspectives on Machine Learning

Published by
Telecommunications Association Inc.

ISSN 2203-1693

The *Journal of Telecommunications and the Digital Economy* is published by TelSoc four times a year, in March, June, September and December.

Journal of Telecommunications and the Digital Economy

Volume 12, Number 3
September 2024

Table of Contents

History of Telecommunications

Editorial Team

The *Journal* is published by the Telecommunications Association (TelSoc), a not-for-profit society registered as an incorporated association. It is the Australian telecommunication industry's oldest learned society. The *Journal* has been published (with various titles) since 1935.

Editorial

Operational AI

Leith H. Campbell
Managing Editor

Abstract: This editorial provides some reasoning behind the decision to publish the Special Issue section on "Perspectives on Machine Learning". It also outlines the other content of this September issue and introduces the new Managing Editor of the *Journal* to take over in 2025.

Keywords: Editorial, Machine Learning, Managing Editor

Artificial Intelligence and Machine Learning

Over the past few years, the *Journal* has built relationships with several conferences – to identify new and interesting areas of research on telecommunications and the digital economy; and as a source of good submissions. A significant number of papers – usually extensions of work originally presented at a conference – have been published as a result.

In looking at the stream of potential papers coming from these conferences, it was clear that there were many research groups using machine learning as a technique. With a dataset of suitable size from some application, one can train a machine-learning model and then test its predictive power on unseen data. Essentially, this is just a new and powerful way to identify important characteristics in a large dataset and to build, automatically, a predictive model. It replaces, or often incorporates, other statistical packages and procedures that have been used in the past. The automated production of a predictive model is often an attractive feature.

This is Artificial Intelligence (AI) being used operationally. While there has been much public discussion — some of it in the nature of a moral panic — about the potential benefits or harms associated with AI, this is the less scary end of the subject — no deepfakes or disinformation. The techniques of machine learning are already embedded in many support systems and research is identifying new potential areas of application.

https://doi.org/10.18080/jtde.v12n3.1046

It seemed timely, therefore, to provide some further data for the discussion of AI by bringing together a collection of papers that apply an AI technique — machine learning — to a variety of applications. Thus, the idea for a special issue on "Perspectives on Machine Learning" was born.

In seeking a guest editor, we approached several university staff working in relevant areas. Professor Haw Su-Cheng from Multimedia University in Malaysia responded enthusiastically and took up the challenge to curate a special issue. She invited papers from a range of university groups: 8 submissions were received and, after the usual peer review, 6 have been accepted for publication. I would like to thank Professor Haw for all her efforts in making the special issue possible.

Professor Haw's introductory editorial (Haw, 2024) is the first paper in this issue.

In This Issue

The Special-Issue section on "Perspectives on Machine Learning", with an editorial and 6 papers, makes up the bulk of the September issue. However, we do also include other papers.

In the Digital Economy & Society section, there are two papers. *Exploring Digital Proficiency Among Mothers of School-going Children in Kerala* uses a survey of 1,000 homemakers in India about their supervision and guidance of their children's use of digital services to make recommendations about how to support this important task. *How Does the Digital Economy Influence the Pursuit of Sustainable Development Goals?* analyses data from both developed and developing countries to identify how the digital economy affects sustainable development.

We also publish a Review paper, *Developing a Holistic Framework for Assessing the Development of the Digital Economy: A Systematic Review of Key Dimensions and Indicators*, which reviews the academic literature and identifies four core dimensions for assessing the development of the digital economy.

As a complement to our Special Issue section, in the History of Telecommunications section, we reprint a paper from 1970 on *An Early Management System for Subscriber Equipment Installation*. No machine learning included – instead, a hand-crafted workflow process.

As always, we encourage you to consider submitting articles to the *Journal* and we welcome comments and suggestions on which topics or special issues would be of interest. Feedback on the current issue would be welcome.

A New Managing Editor for 2025

I will be retiring from the role of Managing Editor at the end of volume 12 in December 2024. A new Managing Editor, Dr Michael de Percy, will take over from January 2025. Dr de Percy was appointed by the TelSoc Board after consideration of the candidates who expressed interest in the position.

The *Journal* has benefitted from Michael de Percy's support for the past 8 years, first as Section Editor, Telecommunications and, for the past 3 years, as Section Editor, Public Policy. He is also a published author in the *Journal*.

Michael de Percy is Senior Lecturer in Political Science at the University of Canberra's School of Politics, Economics and Society. He is also a weekly columnist for *The Spectator Australia*. He holds a PhD in Political Science from the Australian National University and is currently a member of the Australian Research Council's College of Experts. His research background has been in public policy on telecommunications and transport; and he has been a public commentator on the development of the Australian Government's National Broadband Network. Michael is currently the Vice-President of TelSoc, the publisher of the *Journal*.

We can look forward to a smooth transition to the new Managing Editor in the new year and a refreshed approach to the further development of the *Journal*.

Reference

Haw, S.-C. (2024). Editorial: Perspectives on Machine Learning. *Journal of Telecommunications and the Digital Economy*, *12*(3), 1–6. https://doi.org/10.18080/v12n3.1042

Editorial

Perspectives on Machine Learning

Su-Cheng Haw
Multimedia University, Malaysia

Abstract: Progress in machine learning technology has truly impacted our lives by tailoring many of our daily experiences to be seamless and intuitive. This innovation has brought about changes in day-to-day routines; from suggesting music based on our emotions to offering recommendations for places to visit or meals to try out. This special issue explores various Machine Learning technologies. Among some are Machine Learning advances that improve human interaction, predict user behaviours, analyse user reviews, and optimize high-risk investments like Bitcoin trading. These technologies enhance user experiences, help businesses refine marketing strategies, and provide quick insights from vast amounts of information, elevating AI to new heights. With the rise of transformation into advanced technologies taking prominence in our lives, we expect to see these machine learning innovations being integrated across many sectors and uses.

Keywords: Editorial, Machine Learning.

Designing the Present, Creating the Future with Machine Learning

Machine Learning (ML) is a more specialized area of artificial intelligence (AI), which enables systems to learn from data and enhance their performance over time without being explicitly programmed (Catterwell, 2020; Kufel *et al.*, 2023). On the other hand, AI comprises a broader set of technologies targeted at mimicking human intelligence that includes reasoning, decision-making and problem-solving. The focus of ML lies in pattern recognition, prediction and repetitive tasks that are closer to our day-to-day lives (Ali *et al.*, 2024; Mandalapu *et al.*, 2023; Noriega *et al.*, 2023).

In recent years, the advance of machine-learning technology has greatly enriched human life. It simplifies many things we once thought difficult and makes many day-to-day things smarter

https://doi.org/10.18080/jtde.v12n3.1042

and more efficient. End-user applications such as fraud detection (Khalid *et al.*, 2024; Prasad *et al.*, 2023), chatbots, and product recommendations are second nature to us today, making it easier for human-to-technology interaction. Be it suggesting a song that matches our mood (Garanayak *et al.*, 2022; Gatta *et al.*, 2023) or even assisting in deciding where to head for lunch (Dhiman *et al.*, 2024; Lambay & Pakkir Mohideen, 2022), such seamless personalisation of experience comes through ML. Navigation apps, such as Google Maps, provide us with the best and shortest route and real-time traffic information to avoid delays, so that you may get to your destination faster (Yatnalkar *et al.*, 2020). In the context of social media, by learning about your interactions with content, social media platforms are able to find more relevant posts to recommend to you. For all these applications, ML technology makes them a little bit easier or more intuitive. These advances, unimaginable just a few years ago, could profoundly improve lives.

In essence, ML is like having a helpful assistant working behind the scenes (Gopalakrishnan & Kumaran, 2022), continuously learning more about who we are and what we need. It is critical to continue understanding this technology's functions and effects as it develops. ML enables us to benefit from its advantages while making sure that any potential drawbacks, such as biases or privacy issues, are addressed.

In This Issue

Six major areas are covered in this issue, ranging from book recommendations to human interaction to text summarization.

The work by Kumaresan *et al.* (2024) is significant because it describes the abilities of machines to recognise and interpret human postures. Precise estimation of human poses in computer vision is essential for many uses, including interactive gaming, health monitoring, and even safer self-driving cars. The authors present a novel approach to improve Human Pose Estimation (HPE) in computer vision using dual self-attention (DSA) mechanisms embedded in a high-resolution network (HR Net). The DSA mechanisms, which increase global perception, are embedded into HR Net to form a lightweight and powerful architecture, DSA-HR Net. The model is evaluated on a context validation dataset. The improvements in accuracy demonstrate that this model is not only faster but also more accurate than alternative approaches. This could result in pose identification that is more accurate in real-world applications, improving the intelligence and responsiveness of technology to human movements.

The significance of the study by Raja Sekaran *et al.* (2024) lies in its ability to improve smartphone comprehension and interpretation of human behaviours by utilising data from several sensors. Using information from sensors such as accelerometers and gyroscopes, Human Activity Recognition (HAR) aims to allow devices to identify a person's activity accurately, be it walking, running, or even just standing motionless. In order to extract the features of every sensor independently, the authors proposed a lightweight deep temporal learning model called Feature-Level Fusion Multi-Sensor Aggregation Temporal Network (FLF-MSATN). FLF-MSATN can achieve better recognition performance by concatenating more extended temporal information. This method is very important because it allows our everyday gadgets to identify tasks precisely as well as reliably. Maybe your smartphone will become an even better fitness tracker, helping you to keep healthy in more comprehensive ways and reducing risks of falls by the elderly or those in care.

Chongwarin *et al.* (2024) have proposed a new approach to finding user preference trends by analysing a million user reviews to examine the effects of several methods (such as k-Nearest Neighbours and Matrix Factorisation). Their collaborative filtering approach suggests books based on mutual interests. They highlight that it is crucial to match the right algorithm to the features of the dataset, because this interaction has a big impact on how effective the recommendation system can be. In the end, their study aids in the creation of more accurate and customised book recommendations, improving the search experience by assisting users in finding books that immediately and easily correspond with their interests.

The study by Lau & Tan (2024) is crucial because it enhances the automated summary of long documents, and therefore it allows us to quickly read through vast amounts of text. The proposed approach uses an extractive method to complement the abstractive approach of long-form text summarisation generation rephrasing, by finding and selecting important sentences to construct a summary which is more humanly readable. In this way, it ensures that the summary does not lose factual information from the original sources. This method addresses the issues associated with solely relying on human evaluations, and promises a text summarization solution that is more accurate and efficient.

The study by Wong *et al.* (2024) is significant because it supports businesses in developing more accurate marketing strategies, by assisting them to understand and better target their customers. Using Hierarchical Clustering with Recency, Frequency, and Monetary (RFM) Analysis, the authors can segment customer data more accurately as compared to other conventional ways, such as K-Means Clustering. Thus, businesses can more accurately direct their marketing campaigns and ensure the highest accuracy of delivering the right message to the right customer. This can help businesses to achieve higher customer satisfaction, greater customer loyalty and eventually higher revenue.

The new technique proposed by Tay & Lim (2024) could enhance high-risk bitcoin trading investment based on Deep Reinforcement Learning (DRL) algorithms to predict the future

https://doi.org/10.18080/jtde.v12n3.1042

prices of well-known cryptocurrencies, such as DogeCoin, Ethereum, and Bitcoin. This methodology gives hope to investors that they will somehow find a way to increase their wealth even as inflation steadily depletes savings.

Conclusion

This special issue delves into the topic of "Perspectives on Machine Learning", examining diverse perspectives that showcase how ML is reshaping different sectors and driving technological progress that positively and significantly can impact on many areas of life. The perspectives highlighted in this special issue lay the groundwork for understanding ML. We hope they will provide our readers with a deeper insight into the latest developments, insights and the prospective path of this revolutionary domain. It is our hope that this issue will inform and inspire, equipping many readers with the knowledge to navigate and contribute to the evolving landscape of ML.

While ML offers many benefits, such as efficiency, personalisation, and insights, it is essential to balance its advantages with potential limitations. For instance, ML may not be useful in areas that requires human empathy and ethical decision-making. Privacy is another area where misuse or unintended use of personal data can have serious risks. Over-reliance on machine learning in these areas could create dehumanisation and ethical challenges.

References

Ali, Y., Hussain, F., & Haque, M. M. (2024). Advances, challenges, and future research needs in machine learning-based crash prediction models: A systematic review. *Accident Analysis and Prevention*, *194*, 107378. https://doi.org/10.1016/j.aap.2023.107378

Catterwell, R. (2020). Automation in contract interpretation*. *Law, Innovation and Technology*, *12*(1), 81–112. https://doi.org/10.1080/17579961.2020.1727068

Chongwarin, J., Manorom, P., Chaichuay, V., Boongoen, T., Li, C., & Chansanam, W. (2024). Enhancing Book Recommendation Accuracy through User Rating Analysis and Collaborative Filtering Techniques. *Journal of Telecommunications and the Digital Economy*, *12*(3), 51–72. https://doi.org/10.18080/v12n3.976

Dhiman, G., Gupta, G., & Sidhu, B. K. (2024). A Review and Research Panorama on Food Recommender System Based on Health Care. *International Journal of Intelligent Systems and Applications in Engineering*, *12*(17s), 409–422. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/4890

Garanayak, M., Nayak, S. K., Sangeetha, K., Choudhury, T., & Shitharth, S. (2022). Content and Popularity-Based Music Recommendation System. *International Journal of Information System Modeling and Design*, *13*(7), 1–14. https://doi.org/10.4018/ijismd.315027

Gatta, V. La, Moscato, V., Pennone, M., Postiglione, M., & Sperli, G. (2023). Music Recommendation via Hypergraph Embedding. *IEEE Transactions on Neural*

Networks and Learning Systems, *34*(10), 7887–7899. https://doi.org/10.1109/TNNLS.2022.3146968

Gopalakrishnan, S., & Kumaran, M. S. (2022). IIOT framework based ML model to improve automobile industry product. *Intelligent Automation and Soft Computing*, *31*(3), 1435–1449. https://doi.org/10.32604/IASC.2022.020660

Khalid, A. R., Owoh, N., Uthmani, O., Ashawa, M., Osamor, J., & Adejoh, J. (2024). Enhancing Credit Card Fraud Detection: An Ensemble Machine Learning Approach. *Big Data and Cognitive Computing*, *8*(1), 6. https://doi.org/10.3390/bdcc8010006

Kufel, J., Bargieł-Łączek, K., Kocot, S., Koźlik, M., Bartnikowska, W., Janik, M., Czogalik, Ł., Dudek, P., Magiera, M., Lis, A., Paszkiewicz, I., Nawrat, Z., Cebula, M., & Gruszczyńska, K. (2023). What Is Machine Learning, Artificial Neural Networks and Deep Learning?—Examples of Practical Applications in Medicine. *Diagnostics*, *13*(15), 2582. https://doi.org/10.3390/diagnostics13152582

Kumaresan, S. P., Yee, L. F., Palanichamy, N., Annan, E. (2024). Improving Human Pose Estimation with Integrated Dual Self-Attention Mechanism in High-Resolution Network. *Journal of Telecommunications and the Digital Economy*, *12*(3), 7–28. https://doi.org/10.18080/v12n3.984

Lambay, M. A., & Pakkir Mohideen, S. (2022). A Hybrid Approach Based Diet Recommendation System Using ML and Big Data Analytics. *Journal of Mobile Multimedia*, *18*(6), 1541–1560. https://doi.org/10.13052/jmm1550-4646.1864

Lau, A. J. J., & Tan, C. W. (2024). LongT5Rank: A Novel Integrated Hybrid Approach for Text Summarisation. *Journal of Telecommunications and the Digital Economy*, *12*(3), 73–96. https://doi.org/10.18080/v12n3.977

Mandalapu, V., Elluri, L., Vyas, P., & Roy, N. (2023). Crime Prediction Using Machine Learning and Deep Learning: A Systematic Review and Future Directions. *IEEE Access*, *11*, 60153–60170. https://doi.org/10.1109/ACCESS.2023.3286344

Noriega, J. P., Rivera, L. A., & Herrera, J. A. (2023). Machine Learning for Credit Risk Prediction: A Systematic Literature Review. *Data*, *8*(11). https://doi.org/10.3390/data8110169

Prasad, P. Y., Chowdarv, A. S., Bavitha, C., Mounisha, E., & Reethika, C. (2023). A Comparison Study of Fraud Detection in Usage of Credit Cards using Machine Learning. *7th International Conference on Trends in Electronics and Informatics, ICOEI 2023—Proceedings*. https://doi.org/10.1109/ICOEI56765.2023.10125838

Raja Sekaran, S., Han, P. Y., You, L. Z., & Yin, S. H. (2024). Feature-Level Fusion Multi-Sensor Aggregation Temporal Network for Smartphone-Based Human Activity Recognition. *Journal of Telecommunications and the Digital Economy*, *12*(3), 29–50. https://doi.org/10.18080/v12n3.979

Tay, X. H., & Lim, S. M. (2024). Deep Reinforcement Learning in Cryptocurrency Trading: A Profitable Approach. *Journal of Telecommunications and the Digital Economy*, *12*(3), 126–147. https://doi.org/10.18080/v12n3.985

Wong, C. G., Tong, G. K., & Haw, S. C. (2024). Exploring Customer Segmentation in E-Commerce using RFM Analysis with Clustering Techniques. *Journal of*

Telecommunications and the Digital Economy, *12*(3), 97–125. https://doi.org/10.18080/v12n3.978

Yatnalkar, G., Narman, H. S., & Malik, H. (2020). An Enhanced Ride Sharing Model Based on Human Characteristics and Machine Learning Recommender System. *Procedia Computer Science*, *170*, 626–633. https://doi.org/10.1016/j.procs.2020.03.135

Improving Human Pose Estimation with Integrated Dual Self-Attention Mechanism in High-Resolution Network

S Prabha Kumaresan
Faculty of Computing and Informatics, Multimedia University, Cyberjaya, Malaysia

Lee Fong Yee
Faculty of Computing and Informatics, Multimedia University, Cyberjaya, Malaysia

Naveen Palanichamy
Faculty of Computing and Informatics, Multimedia University, Cyberjaya, Malaysia

Elham Annan
Faculty of Computing and Informatics, Multimedia University, Cyberjaya, Malaysia

Abstract: Human Pose Estimation (HPE) in computer vision (CV) has garnered significant attention due to its diverse applications. Deep convolutional neural networks (CNNs) may be solutions for addressing this challenge, but still face several critical issues. Many existing models employ serial convolution with pooling, leading to low-resolution outputs that are suboptimal for the precise localisation required in HPE. They often prioritise local feature learning, overlooking crucial contextual relationships between key-points. This work addresses these challenges by proposing a novel approach for enhancing HPE. Firstly, the paper evaluates the high-resolution network (HRNet) and its comparative advantages over other CNN architectures. Secondly, it introduces a dual self-attention (DSA) mechanism designed to enhance the model's global awareness, thereby enriching feature maps with contextual information. Thirdly, it integrates the DSA mechanism into HRNet, crafting DSA-HRNet. The model performance was tested on the COCO Val 2017 validation dataset, showing improvements of 2.3% in mean average precision (mAP), 3% in AP at 50 (AP50), and 2.7% in AP at 75 (AP75). Finally, the work includes an investigation into the effectiveness of the DSA mechanism within the HRNet framework, through a series of experiments, showing this work offers a streamlined and effective solution for improving HPE.

https://doi.org/10.18080/jtde.v12n3.984

Keywords: Computer Vision, Human Pose Estimation, Convolution Neural Network, High Resolution Network, Dual Self-Attention.

Introduction

Computer vision (CV), a cornerstone of Artificial Intelligence (AI), has emerged as a focal point of global interest. Its core concept revolves around leveraging various AI algorithms to analyse input data and extract meaningful insights (Voulodimos *et al*., 2018). Within the realm of CV, human pose estimation (HPE) represents a specific challenge sharing the foundational principles of CV. HPE aims to construct human skeletons by detecting key-points on human instances through algorithmic processes applied to input data. The efficacy of HPE models, often measured by metrics like average precision (AP), heavily depends on the precision of key-point localisation achieved by these algorithms. While traditional methods, such as histogram of oriented gradients (HOG) and pictorial structure model (PSM), have been employed, they tend to be instance-based and have limited generalisation capabilities. In contrast, convolutional neural networks (CNNs) have shown immense promise in addressing HPE challenges. CNNs not only possess strong generalisation abilities but also excel in extracting more meaningful features compared to traditional methods. With the evolution of input complexity and resolution, CNN architectures have advanced into deep neural networks (DNNs), incorporating more intricate convolutions and operations, which represent the contemporary trend in addressing complex HPE challenges.

Recent years have witnessed the emergence of various DNN architectures tailored for HPE tasks. A notable development is the utilisation of high-resolution network (HRNet) implemented by Sun *et al.* (2020), which processes HR representations throughout the entire process, demonstrating exceptional performance, particularly in scenarios involving complex inputs, such as crowded scenes with occlusions. Despite significant advancements in DNN architectures for HPE, several challenges persist. Existing architectures may exhibit deficiencies in feature learning, focusing excessively on local features while overlooking global dependencies. The prevalent serial convolution learning paradigm may lead to information loss for position-sensitive tasks like HPE. Attention mechanisms, while enhancing model performance, often incur a substantial increase in computational overhead due to their repetitive integration within models. The shortcomings of traditional DNNs for HPE include the generation of low-resolution representations, neglect of global dependencies during convolution, and increased computational requirements for attention mechanisms.

https://doi.org/10.18080/jtde.v12n3.984

Prior works

In the realm of two dimensional (2D) HPE (2D-HPE), significant strides have been made over the years, with various methodologies and models contributing to its advancement. Deep Pose, a regression-based method for single-person pose estimation was developed by Toshev & Szegedy (2014). It utilises a DNN architecture to directly predict key-point coordinates from RGB images. Although effective, it lacks consideration for the inner relationships between key-points. Meanwhile, attention mechanisms have been proposed in natural language processing (NLP) to emphasize relevant information during translation tasks (Bahdanau *et al.*, 2014). This concept was later adapted to CV, aiding in the focus on important regions in feature maps. Furthermore, to improve precision in key-point prediction, traditional DNN methods have been modified by adding an iterative feedback mechanism in Carreira *et al.* (2016). This method enhances initial solutions using error prediction results but still lacks consideration for key-point relationships. A mask region-based CNN (R-CNN) was developed by He *et al.* (2017), which fine-tunes faster R-CNN by adding an additional object mask to improve multi-person pose estimation. Although the model achieves better performance compared to previous approaches, it still struggles with complex poses and occlusions.

Self-attention (SA) has become a key component of transformers, allowing the model to capture dependencies between input elements effectively. It has been successfully applied to various tasks beyond sequence-to-sequence tasks, including computer vision tasks, due to its ability to capture long-range dependencies and relationships within data (Vaswami *et al.*, 2017). On the other hand, the two-stage model proposed by Papandreou *et al.* (2017), faces challenges in handling complex poses and occlusions, as it relies on a sequential process of detection and prediction. Additionally, this sequential approach leads to error propagation, especially when detecting multiple persons with varying poses and occlusions. Open-pose, developed by Cao *et al.* (2017), demonstrates superior runtime speed and accuracy. However, it struggles with accuracy in challenging scenarios, such as crowded scenes or extreme poses. Open-pose's reliance on part affinity fields (PAFs) for key-point association leads to inaccuracies in cases of occlusion or overlapping body parts. A cascaded pyramid network (CPN) was proposed by Chen *et al.* (2018), demonstrating excellent performance in handling occlusion problems. However, it suffers from increased computational complexity due to its cascaded architecture. Additionally, CPN requires significant computational resources during training and inference, limiting its practical applicability in real-time or resource-constrained scenarios. Following that, heatmap-based methods have been extended to video sequence processing by Luo *et al.* (2018), using long short term memory (LSTM), which introduces additional computational complexity and potential challenges in capturing long-term temporal dependencies. Moreover, LSTM-based approaches struggle with real-time

https://doi.org/10.18080/jtde.v12n3.984

performance and require optimisation for efficient inference on resource-constrained devices. Meanwhile, HRNet, which leverages parallel convolution and multi-scale fusion to maintain high-resolution representations and exchange information across different resolutions, was introduced by Sun *et al.* (2019). Even though HRNet demonstrates superior performance in pose estimation tasks, it requires significant computational resources.

To overcome these limitations, a dual SA (DSA) mechanism has been proposed in Fu *et al.* (2019), which combines spatial, channel, and self-attention mechanisms. DSA enhances feature representation by considering both spatial and channel-wise dependencies, aiding in tasks like image segmentation. Furthermore, vision transformers (VTs) were introduced by Dosovitskiy *et al.* (2020), which applies the transformer architecture directly to image data. VT divides images into patches, applies self-attention to capture global dependencies, and achieves competitive performance in image classification tasks. Following that, to enhance feature representation, HRNet was modified by Li *et al.* (2020), by introducing additional channel convolutions, achieving improved performance but at the cost of increased model complexity.

While unipose extends heatmap-based methods to video sequence processing, it faces challenges in handling temporal variations and occlusions in dynamic scenes. The reliance on encoder-decoder architectures and LSTM may introduce computational overhead and potential difficulties in capturing long-term dependencies in video sequences (Artacho & Savakis, 2020). To improve pose estimation performance, an attention refined network (AR Net) was proposed by Wang *et al.* (2021), adding a refined attention module to the output layer of HRNet, capturing spatial and channel-wise attention information. Furthermore, HRFormer has been introduced in Yuan *et al.* (2021), integrating the parallel convolution paradigm of HRNet with transformer-based self-attention. HRFormer aggregates feature map patches using multi-head attention, enhancing feature representation for pose estimation. The coordinate attention module was introduced by Wu *et al.* (2023), extracting features from both width and height directions to construct attention maps. This module enhances HRNet's feature representation by considering spatial relationships. Building on the advances in feature extraction and deep learning, Ti *et al.* (2023) introduced a deep learning network trained to recognize gender from gait features. This method is effective because deep learning can find complex patterns in the gait data. Tests on two datasets, one called "Gait in the Wild" and another self-collected one, show that GenReGait is reliable and accurate in recognizing gender. In a related study, Tan *et al.* (2024) developed a method to analyse how people walk using videos from cameras without any markers on the body. Videos were collected from two angles (side and front) using smartphones. Additionally, different aspects of walking were measured based on the joint positions.

Motivation and contribution

Based on the above discussion, this paper focuses on enhancing an existing network backbone to improve performance in 2D-HPE tasks. It presents a novel approach leveraging the HRNet and introduces a DSA-HRNet to address key challenges in feature learning and attention mechanisms. The contribution of this work includes:

1. Introducing the HRNet as a backbone for pose estimation;
2. Incorporating a parallel convolution learning paradigm to maintain HR representations and improve prediction accuracy;
3. Implementing a DSA mechanism to capture global dependencies and contextual information within the HRNet architecture, thereby enhancing model performance without a disproportionate increase in computational complexity. Finally empirical evaluations and experiments are conducted to validate the effectiveness of the proposed enhancements and their impact on HPE performance.

Research Methodology and Data Preparation

This paper aims to enhance the performance of existing models by integrating attention mechanisms, specifically focusing on self-attention, inspired by its frequent utilisation in contemporary research, particularly in networks such as HRNet and others.

In previous discussions, it was noted that conventional DNN backbones predominantly employ the serial convolution paradigm, whereas HRNet introduces a novel approach known as parallel convolution to preserve HR representations. Despite its strengths, HRNet lacks the ability to capture contextual information, prompting the exploration of attention mechanisms as a potential remedy for this limitation. This paper proposes the integration of a DSA mechanism with the original HRNet architecture to address the challenge of HPE.

The resulting network, termed DSA-HRNet, leverages the inherent advantages of HRNet while augmenting it with DSA mechanisms to enhance feature extraction and uncover underlying relationships within feature maps. This section commences with an overview of the dataset utilised in this study, followed by detailed exposition of the DSA-HRNet architecture. Additionally, it includes an analysis of performance metrics and describes various experiments conducted, along with their corresponding hyperparameters.

Dataset

This paper examines the efficacy of DSA-HRNet on the common objects in context (COCO) 2017 dataset (Lin *et al.*, 2014), which comprises three subsets: training, validation, and testing. DSA-HRNet is trained on the COCO train2017 subset (118K images) and assesses its

https://doi.org/10.18080/jtde.v12n3.984

performance using COCO Val 2017 (5K images). Given the versatility of the COCO dataset across various tasks, it encompasses diverse annotations. Specifically, for pose estimation, COCO records the coordinates of 17 key-points, as depicted in Figure 1. These key-points are instance-specific and can vary among individuals. To reconstruct the human skeleton, the predicted key-points are linked together using modelling techniques. Figure 1 illustrates the process of human skeleton formation as outlined in this paper.

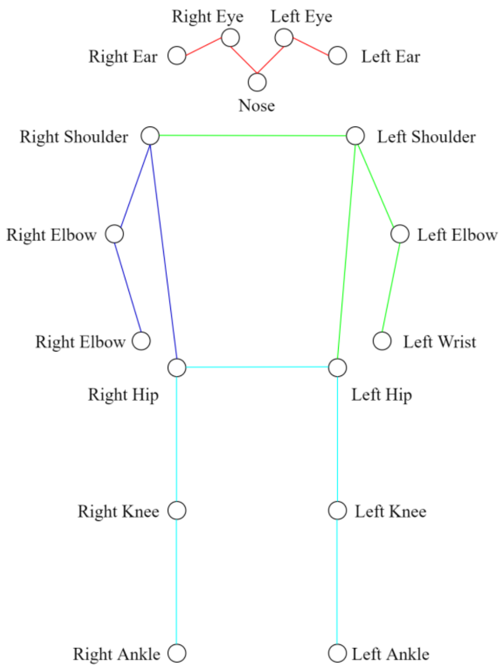


Figure 1. Key-point data captured by the COCO dataset alongside its associated skeleton representation

In terms of occurrences, COCO includes a wide range of human body types, as depicted in Figure 2.

Figure 2. Example images from COCO2017 training set

The crucial steps required for the dataset to ensure its quality and suitability are illustrated in this section.

1) **Data Preprocessing:** As mentioned in Sun *et al.* (2019), to match the input size of DSA-HRNet, all the images are resized into 256x192 (4:3) (height (H) by width (W)) in resolution. At the same time, the boundary of images is extended to avoid information loss. COCO2017 contains images without human instances. Therefore, these images are filtered out after examining their labels. The official COCO implementation contains annotation files for all the images, including information

such as their labels, positions of the key-points etc. These annotation files will be used for the preprocessing of the dataset.

2) **Data Augmentation:** To enhance the model, several data augmentation techniques are employed for HRNet. This paper introduces four types of data augmentation techniques used on training data.
   (i) Half Body: In the COCO2017 dataset, the key-points are grouped based on their location on the human body. The half-body technique only selects the key-points on the upper or lower body parts for estimation. This technique enhances the ability to process images with incomplete key-points.
   (ii) Affine Transform: Affine transformation is a type of transformation that is usually applied in the mathematical domain. Here, the affine transform applied will perform scaling and rotation on the given images. This allows the model to capture information about human instances from different angles.
   (iii) Random horizontal Flip: This type of flipping augments the dataset by applying a random flip on a given image, resulting in the mirror image of the original version.
   (iv) Key-point to heatmap: The final type of data augmentation will enhance the quality of key-points by transforming them into a heatmap. This is achieved by applying a Gaussian kernel to the original images.

Model instantiation

In this section, the components that constitute the backbone of the HR model and the supplementary modules, such as the stem, DSA module and heatmap regressor, are explored, detailing each module's functionality and contribution towards the instantiation of the model. The stem is used for reducing the resolution of the original input, while the heatmap regressor is used to generate heatmaps for key-point prediction.

Stem

The stem, functioning as the preprocessing step, reduces the resolution of the original input using two 3x3 kernels with a stride of 2 and a channel (C) size of 64 for input convolution. The processed feature maps are then further utilized by the HR module.

High Resolution module

This paper utilized HRNet-W32 with four stages introduced by Wang *et al.* (2020), as the backbone. Its architecture is the same as Sun *et al.* (2019), except for a final fourth stage with a newly introduced parallel convolution.

https://doi.org/10.18080/jtde.v12n3.984

To better illustrate it, this paper introduces a different type of notation to represent HRNet-W32. It is like that in Sun *et al.* (2019), yet with an additional superscript *c*, becoming $\mathcal{N}_{sr}^{c}$ to display the channel number in addition to subnetwork $\mathcal{N}$ in the *s-th* stage and *r* the resolution index. Hence, the following annotation can be used to demonstrate HRNet-W32. In HRNet-W32, 32 refers to the base channel of the feature maps, i.e., the channel number used by the feature map with the largest resolution. The network structure, containing 4 parallel sub-networks, is given in (1) as follows,

$$\begin{array}{llll} \mathcal{N}_{11}^{32} \rightarrow \mathcal{N}_{21}^{32} \rightarrow \mathcal{N}_{31}^{32} & \rightarrow \mathcal{N}_{41}^{32} \\ \quad\;\; \searrow \mathcal{N}_{22}^{64} \rightarrow \mathcal{N}_{32}^{64} & \rightarrow \mathcal{N}_{42}^{64} \\ \qquad\qquad\;\; \searrow \mathcal{N}_{33}^{128} & \rightarrow \mathcal{N}_{43}^{128} \\ & \searrow \mathcal{N}_{44}^{256} \end{array} \tag{1}$$

The intricate components, such as bottleneck, transition structure and parallel convolution structure, that constitute the backbone of the HR model are illustrated in this section.

Bottleneck (Stage 1)

When constructing the HRNet, the bottleneck architecture used in ResNet50 is applied. Stage 1 is directly connected to the previous stem. The bottleneck structure is the main component of the first stage. It is used to process the input and increase the depth of the feature maps. Figure 3 demonstrates a single bottleneck.

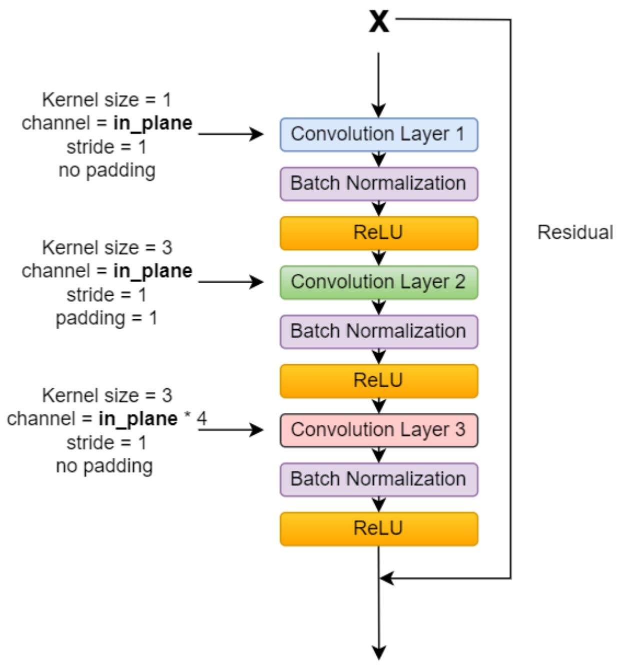


Figure 3. Inner structure of bottleneck

The bottleneck receives a feature map as its input. Within the bottleneck, three convolutions are performed. It is worth noting that there exists a hyperparameter "in_plane" in each bottleneck. It determines the number of channels of feature maps inside the bottleneck. After

https://doi.org/10.18080/jtde.v12n3.984

passing the bottleneck, the number of channels will become four times that of the "in_plane". At stage 1 of the HRNet, the bottleneck block will be repetitively conducted four times. The details of each bottleneck block are summarised in Table 1.

Table 1. Setting of the bottleneck of stage 1

Name	in_plane	output size (CxHxW)
Bottleneck1	64	256 x 64 x 48
Bottleneck2	64	256 x 64 x 48
Bottleneck3	64	256 x 64 x 48
Bottleneck4	64	256 x 64 x 48

After stage 1, the feature map is 64x48x256 and will be used in the later stage.

Transition structure

As highlighted previously, whenever the model enters the next stage, a new branch is added to the network. To achieve such effects, there exists a transition structure between two stages. It is noteworthy that the transition layers between stage 1 and stage 2 are different from the transition layers in the latter stage. Their difference lies in the channel number. Because the output size of the bottleneck is 256 x 64 x 48 and 256 is not the base channel number for the HRNet, there must be a conversion from 256 to 32 (base channel number for HRNet-W32). This operation is achieved in the transition from stage 1 to stage 2. The comprehensive operation is shown in Figure 4(a). After the transition structure, the channel numbers of two branches in stage 2 become 32 and 64. At the same time, the resolution of the feature maps in the new branch becomes 1/2 of the upper branch. For the transition layers in the later stages, the operations are shown in Figure 4(b).

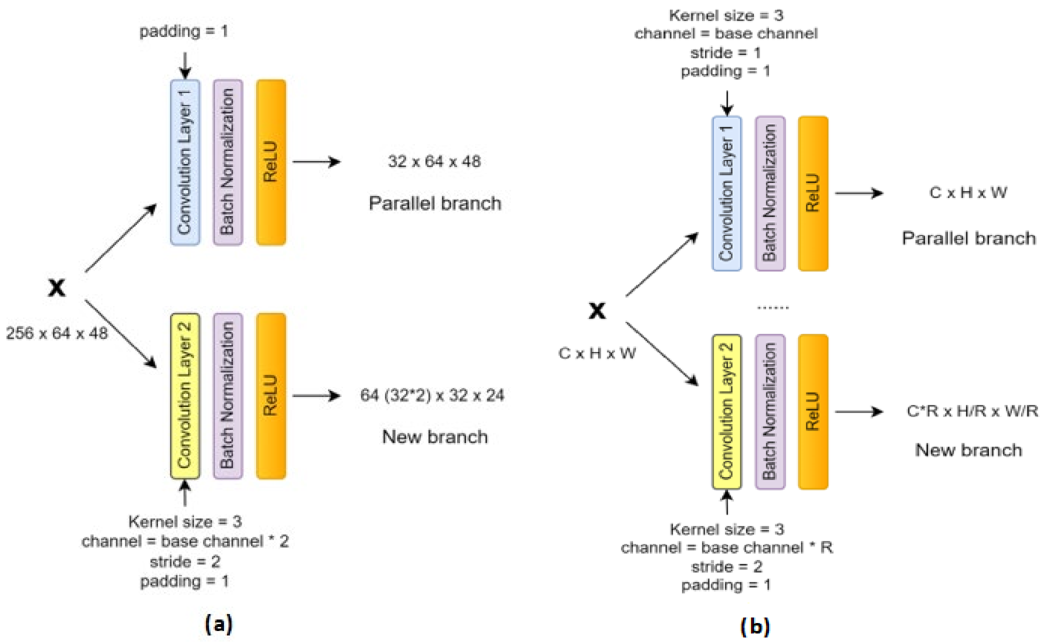


Figure 4. Transition structure of different stages: (a) Transition structure between stage 1 and stage 2; (b) Transition structure between other stages

https://doi.org/10.18080/jtde.v12n3.984

Figure 4(b) introduces another notation R to describe the relationship $R = 2^{stage-1}$ between the largest feature map and the newly generated one. For instance, at the third stage, the resolution of the feature map is 128 x 16 x 12, where the channel is 4*32 and the resolution is 1/4 of the first stage.

Parallel convolution structure (Stages 2–4)

As introduced in the HR module, each stage contains several branches that perform convolutions with different resolutions. Each branch consists of four basic blocks. This setting remains the same for the remaining stages. However, these stages are repeated multiple times in Sun *et al.* (2019), to acquire high-quality feature maps. The details of these subsequent stages are shown in Table 2.

Table 2. Setting of the remaining stages

Stage	Feature map size (C x H x W)	Repeated Times
2	64 x 32 x 24	2
3	128 x 16 x 12	4
4	256 x 8 x 6	3

At the end of the stage, multi-scale fusion is conducted to exchange information across resolutions. Therefore, according to Table 2, eight multi-scale fusions are conducted. However, the last multi-scale fusion only contains up sampling and only the feature map with highest resolution will be generated.

Dual Self Attention (DSA) module

The DSA module for image segmentation tasks, employed as a post-processing component, has been utilised in Fu *et al.* (2019). Additionally, an inner channel number for the key and query in the polarised self-attention (PSA) module is specified, which is determined by dividing the input channel by 8. This parameter, termed "inner channel", governs the processing of spatial information by the position attention module (PAM) and is treated as a hyperparameter in the DSA module, subject to experimentation. In this study, the inner channel number in PAM was adjusted to 1, ensuring that convolution layer 1 yields a mapping of identical size. A comparison between two PAM modules is presented in Figure 5, where Figure 5(a) corresponds to the configuration used by Fu *et al.* (2019) and Figure 5(b) corresponds to the configuration introduced in this paper.

Because the two PAMs have different internal channels, there are also differences in their parameters. The one used in Fu *et al.* (2019), has 6296 parameters, while the one introduced in this paper contain 6836 parameters. Though the number of parameters increases, it may contain more position-related information compared with the old version. This paper incorporates the DSA module at the end of HRNet to process the results from the entire network. These attention feature maps will be passed to the heatmap regressor for processing and key-point connection. The modified DSA module and overall framework of the DSA-HRNet can be seen in Figures 6(a) and 6(b).

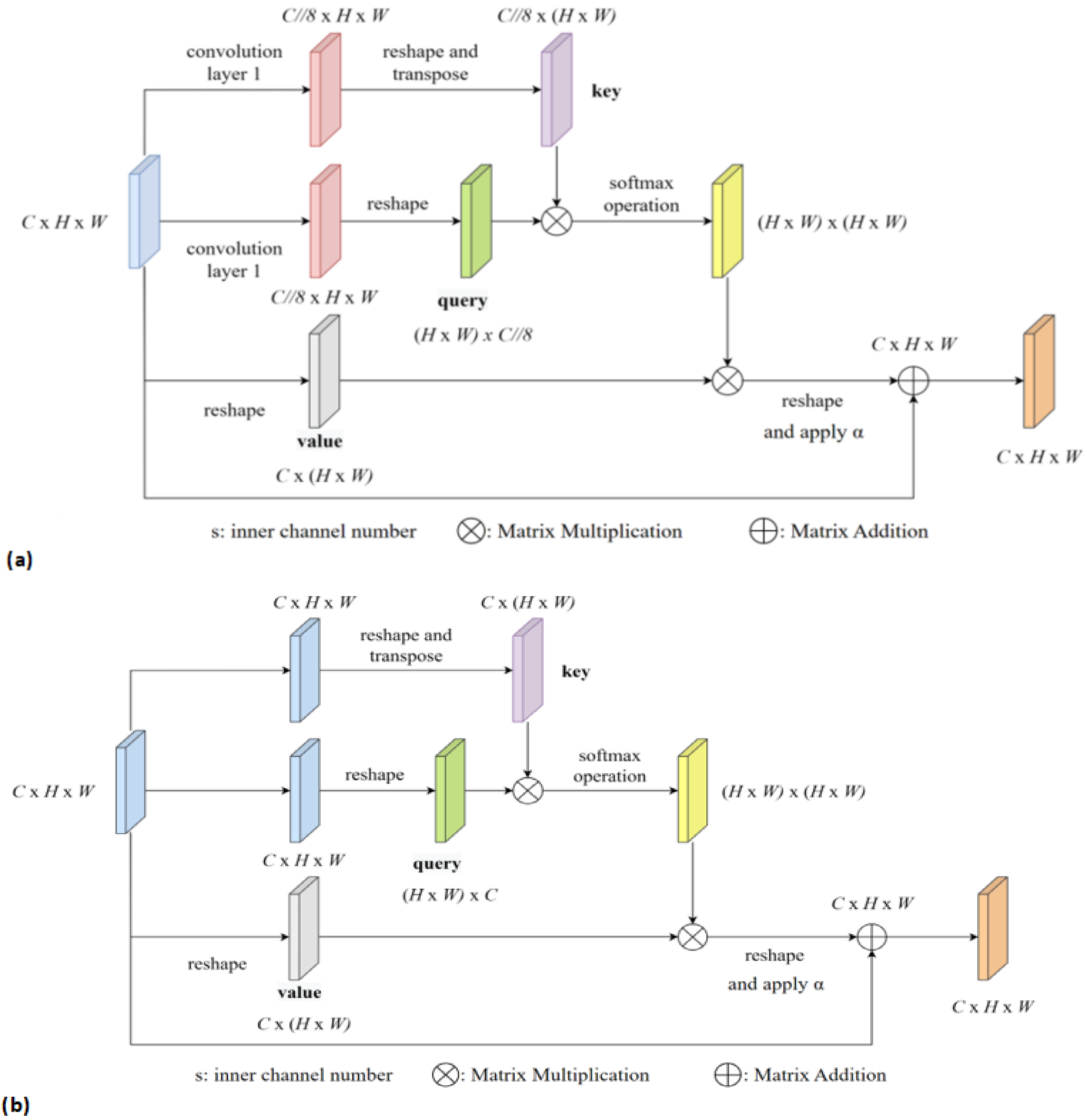


Figure 5. Illustration of two different PAM: (a) PAM used in Fu *et al.* (2019); (b) PAM introduced in this paper

Heatmap Regressor

Heatmap regressor is the last step of the entire HRNet. It receives the feature maps processed by DSA module and encodes them as a heatmap. In other words, the predicted confidence of these key-points will be generated and stacked together into an entire image.

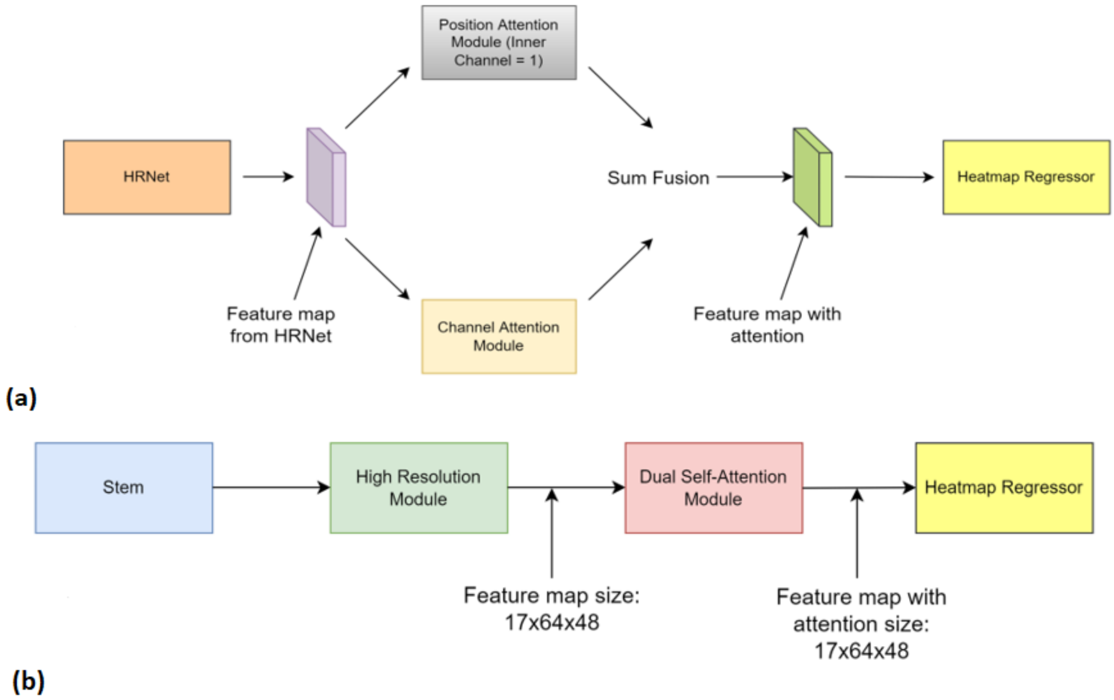


Figure 6. (a) Modified DSA module; (b) Proposed overall model framework

Evaluation metrics

To evaluate the performance of the proposed architecture on the HPE problem, this paper utilizes two evaluation metrics: COCO evaluation metrics; and mean square error (MSE). The COCO evaluation metrics are used to evaluate the performance on COCO Val 2017 at validation stage, while MSE is used during the training process.

COCO evaluation metrics

COCO evaluation metrics are introduced by the COCO dataset (Lin *et al.*, 2014). They are the most commonly used evaluation metrics for the HPE problem. There are two types, comprising ten different evaluation metrics. All these metrics are based on the object key-point similarity (OKS). The OKS can be calculated using (2) where d_i is the Euclidian distance between the i^{th} predicted key-point and its ground truth; k_i is a constant to control falloff for each key-point; v_i is the visibility flag (0 indicates invisible, 1 indicates partially visible, 2 indicates fully visible) and the formula considers its sigmoid distribution δ. Finally, s is the scale factor.

$$\frac{\sum_i exp(-d_i^2/2s^2k_i^2)\delta(v_i>0)}{\sum_i \delta(v_i>0)} \quad (2)$$

The OKS is used to measure the similarity between the predicted key-point and the actual ones, with ranges of 0-1. It is generally combined with AP and Average Recall (AR) and is treated as a threshold. For example, AP50 calculates the AP when OKS = 0.5. The formulas for AP and AR can be seen in (3) and (4), respectively.

$$AP = \frac{TP}{TP+FP} \times 100 \tag{3}$$

$$AR = \frac{TP}{TP+FN} \times 100 \tag{4}$$

Here, the true positive (TP), false positive (FP) and false negative (FN) are determined by the OKS. For example, for OKS = 0.5, when a key-point's OKS > 0.5, it is considered TP; otherwise, it is considered FP. The FN refers to those non-detected or omitted key-points.

Besides the single AP, COCO introduces the mean AP (mAP) and mean AR (mAR). These two metrics consider the AP or AR with 10 different OKS (0.5, 0.55, 0.60, ..., 0.95) and take their average values. Equations (5) and (6) demonstrate their calculation process.

$$mAP = \frac{\sum_{i=10}^{19} AP_{0.05\times i}}{10} \times 100 \tag{5}$$

$$mAR = \frac{\sum_{i=10}^{19} AR_{0.05\times i}}{10} \times 100 \tag{6}$$

In addition, COCO contains the AP and AR for medium and large areas, denoted as AP_M, AP_L, AR_M and AR_L. The 10 COCO evaluation metrics are summarized in Table 3.

Table 3. Interpretation of all 10 COCO evaluation metrics

Name	Interpretation
mAP	mean average precision
AP_{50}	average precision when OKS = 0.5
AP_{75}	average precision when OKS = 0.75
AP_M	average precision for medium area
AP_L	average precision for large area
mAR	mean average recall
AR_{50}	average recall when OKS = 0.5
AR_{75}	average recall when OKS = 0.75
AR_M	average recall for medium area
AR_L	average recall for large area

Mean Square Error

When training the model, MSE is utilised as the loss function (Sun *et al.*, 2019). It first calculates the error of the key-points between the generated heatmap and the actual heatmap. After that, the mean of horizontal and vertical directions is calculated. Finally, it takes the average of the entire batch as the final MSE.

Experimentation and Evaluation

In the previous section, the details of DSA-HRNet have been described. This section mainly introduces all the related experiments with the proposed structure.

Experiment setting

The following methods are used to train the HRNet. First, HRNet is trained on ImageNet and later applied to the HPE problem. The model trained on ImageNet is considered the pretrained model. Since the DSA-HRNet uses the HRNet-W32 as backbone, it can directly employ the best-trained HRNet-W32 and apply the DSA module to it. Compared with Sun *et al.* (2019), the DSA-HRNet does not need to update the parameters of the HRNet part, and the DSA module contains only less than 7k parameters. The number of epochs for training is adjusted accordingly. The training settings are shown in Table 4.

Weight decay (also known as L2 Regularisation) is a method used to adjust the loss during training. Using penalty loss, the parameters are updated. The learning rate decay factor refers to the decay of learning rate. For training DSA-HRNet, the Adaptive Moment Estimation with Weight Decay (ADAM W) optimizer is employed. A multi-step optimiser strategy has been performed for training DSA-HRNet. The model is trained for 210 epochs and, at the 30th and 40th epochs, the learning rate is decreased to 0.0001 and 0.00001, respectively.

Table 4. Training setting of the DSA-HRNet

Optimizer	**AdamW**
Weight decay	0.0001
Learning Rate	0.001
Epoch number	50
Learning rate decay factor	0.1
Batch size	32
workers	8

Experiment 1: Main experiment

The core experiment of this study is to compare the performance of DSA-HRNet with the original pretrained HRNet-W32, with input size 256 x 192. Namely, mAP, AP50, AP75, AP_M, AP_L, and AR will be compared with work referred to in Sun *et al.* (2019). Some other comparison criteria (e.g., the number of parameters) will be considered too. Apart from the original HRNet, some works embedding attention mechanisms into the HRNet will be included in the comparison and the analyses will be performed accordingly.

https://doi.org/10.18080/jtde.v12n3.984

Experiment 2: Variant DSA-HRNet

The proposed DSA-HRNet in this paper only contained one DSA module at the end of the HRNet. In this experiment, to verify the effectiveness of DSA-HRNet, two variants of the DSA-HRNet are proposed. The first one is a stacked version of the DSA-HRNet. That is, one additional DSA module is added after the current DSA module, forming a serial double DSA structure at the end, as seen in Figure 7(a). The later DSA module will process the attention feature maps from the previous DSA module. The other one is the parallel double DSA structure, as seen in Figure 7(b). Similar to the DSA module itself, the feature maps from the HRNet will be processed by two DSA structures simultaneously. After that, the sum fusion of these will be taken and passed to the heatmap regressor.

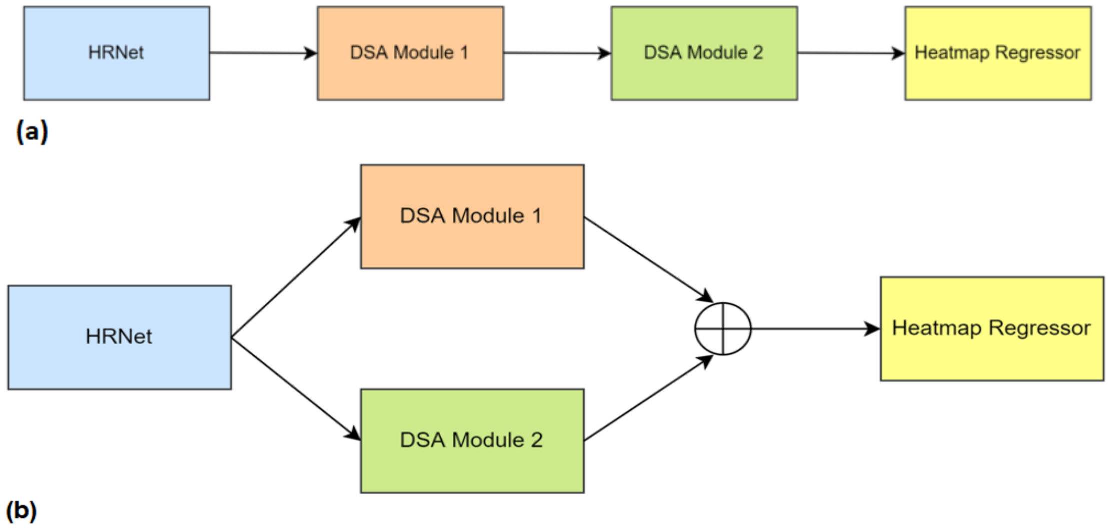


Figure 7. Two Variants of DSA-HRNet: (a) Serial double DSA-HRNet; (b) Parallel double DSA-HRNet

Experiment 3: Hyperparameter testing on DSA-HRNet

Considering the "inner channel" is a hyperparameter in the PAM, it can produce multiple versions of the DSA-HRNet by changing its values. As introduced in Fu *et al.* (2019), the "inner channel" of PAM is set to be 8, whereas, in the DSA-HRNet, the number is set to be 1. To verify the effect of this inner channel, this experiment selects three different values of "inner channel" and tests their performance on the HPE problem. The detailed setting can be seen in Table 5.

Table 5. Inner channel setting of different DSA-HRNet in Experiment 3

Model Name	Inner Channel Number
DSA-HRNet	1
DSA-HRNet_inner_2	2
DSA-HRNet_inner_4	4
DSA-HRNet_inner_8	8

Results and Discussion

This section presents the results from different experiments accompanied by their corresponding comparisons and discussion.

Experiment 1: Main experiment

Figure 8. Training Loss Curve of DSA-HRNet

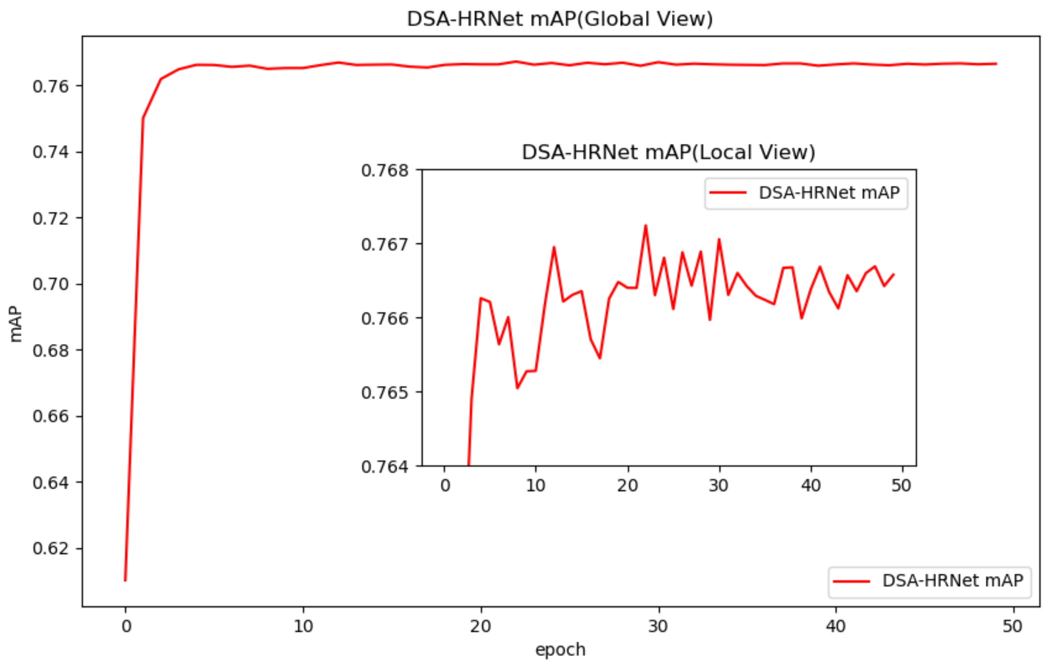


Figure 9. mAP Curve of DSA-HRNet

The training curve and the mAP curve of the DSA-HRNet are presented in Figure 8, where the local view of the curve is contained in a small window. It can be observed that the loss has a

https://doi.org/10.18080/jtde.v12n3.984

huge decrease in the first few epochs. Later, it drops slightly with each epoch, while reaching around 0.0229 in the final epoch.

Considering the mAP curve shown in Figure 9, the DSA-HRNet is built on a pretrained network, and the mAP starts at around 0.6, which is a relatively good result. After a few epochs, the performance rises to around 0.76. The peak value, which is around 0.767, appears at around the 20th epoch. Following that, the mAP fluctuates around 0.766.

Table 6. Comparison between DSA-HRNet and original HRNet on COCO Val 2017

Model	Pretrained	AP	AP_{50}	AP_{75}	AP_M	AP_L	AR
HRNet-W32	No	73.42	89.51	80.73	70.27	80.12	78.93
HRNet-W32	Yes	74.41	90.54	81.92	70.81	81.03	79.84
DSA-HRNet	Yes	76.73	93.55	84.66	74.24	81.15	79.62

The comparison between the DSA-HRNet and the original HRNet is shown in Table 6. Here, this paper utilised the pretrained HRNet-W32 as the model backbone without training the DSA-HRNet from scratch. The results show that the performance of the DSA-HRNet outperforms the original HRNet-W32 on all evaluation metrics except for the AR. The mAP is 2.3% higher than the original HRNet, while the AP50 is 3% better. This indicates that the DSA-HRNet has a better overall performance in predicting key-points on human bodies. In terms of the AR, it is 0.2% less than the pretrained HRNet-W32.

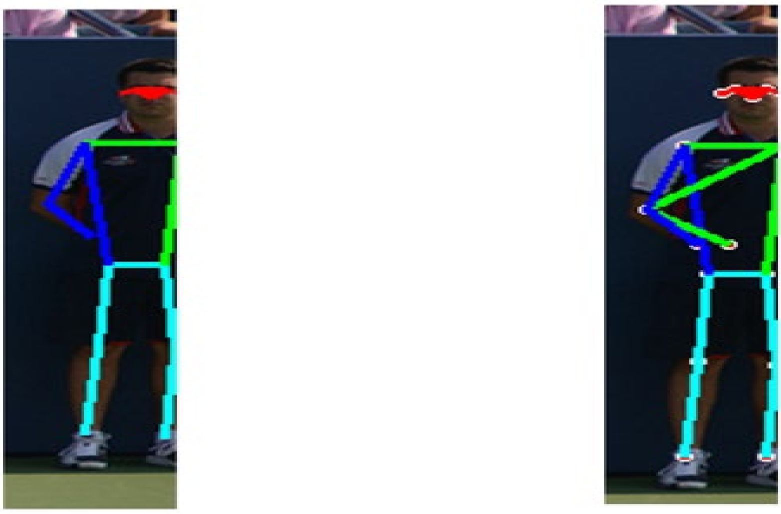
(a) (b)

Figure 10. Pair one — Comparison between (a)DSA-HRNet and (b) HRNet

For better comparison, some of the estimation results from DSA-HRNet and HRNet are presented. In pair one Figure 10(a), it can be observed that, in the DSA-HRNet, the key-point precision has increased a bit compared with the results from Figure 10(b): for example, the wrist and ankle on the predicted human instance. In terms of the second pair, DSA-HRNet provides the correct prediction result, thereby boosting the model's prediction ability, as seen

https://doi.org/10.18080/jtde.v12n3.984

in Figure 11(a). On the other hand, there exists a failed prediction in the HRNet as seen in Figure 11(b).



(a) (b)

Figure 11. Pair two — Comparison between (a)DSA-HRNet and (b) HRNet

In addition to the comparison between the DSA-HRNet and the HRNet-W32, Table 7 displays comparisons between the DSA-HRNet and other attention-based HRNet architectures. This table is ranked based on the mAP values of each model. The DSA-HRNet achieves the highest performance on all the evaluation metrics except for the AR. It shows that the proposed DSA module has the best performance among these architectures. The proposed model framework can also compete with transformer architectures like the HRFormer. Overall, it shows that this proposed method can be a flexible solution for solving the HPE problem.

Table 7. Comparison of DSA-HRNet and other HPE related works on COCO Val 2017

Model	Backbone	Parms	AP	AP_{50}	AP_{75}	AP_M	AP_L	AR
HRFormer-S	Transformer	7.8M	74.0	90.2	81.2	70.4	80.7	79.4
HRNet	HRNet-W32	28.0M	74.4	90.5	81.9	70.8	81.0	79.8
HRFormer-B	Transformer	43.2M	75.6	90.8	82.8	71.7	82.6	80.8
DSA-HRNet	HRNet-W32	28.0M	76.7	93.5	84.6	74.2	81.1	79.6

Experiment 2: Variant DSA-HRNet

The results of these two variant structures can be seen in Table 8 where comparisons of different criteria are made.

From Table 8, it can be seen that the serial double DSA-HRNet obtains the best performance among these three architectures, because of two consecutive attention blocks. In particular, its mAP is roughly 0.13% higher and the AP50 is 0.03% higher than the original network. Since a single DSA module only contains 6836 parameters, it can be inferred that, by stacking multiple DSA modules, the performance of the model can be further improved to a certain degree while keeping the number of parameters at a relatively small scale. In contrast, the parallel one performs approximately the same as the original one except for the AP. This is because the parallel structure utilises two DSA modules to process the feature maps simultaneously without any intermediate convolutions to extract features.

Table 8. Results of two variants of DSA-HRNet on COCO Val 2017

Model	Backbone	Block Parms	AP	AP_{50}	AP_{75}	AP_M	AP_L	AR
DSA-HRNet	HRNet-W32	6836	76.72	93.52	84.57	74.15	81.10	79.59
Serial Double DSA-HRNet	HRNet-W32	13672	76.85	93.55	84.61	74.19	81.21	79.68
Parallel Double DSA-HRNet	HRNet-W32	13672	76.72	93.55	84.56	74.17	81.00	79.54

Experiment 3: Hyperparameter testing on DSA-HRNet

The results of the last experiment are presented in Table 9. As indicated, four different inner channel numbers with their respective results are shown. As their results are quite alike, all the numerical values are rounded to two decimal points to illustrate the details.

Table 9. Results of DSA-HRNet with different inner channel

Model	Inner Channel	Block Parms	AP	AP_{50}	AP_{75}	AP_M	AP_L	AR
DSA-HRNet	1	6836	76.72	93.52	84.57	74.15	81.10	79.59
DSA-HRNet-Inner-2	2	6512	76.69	93.54	84.62	74.29	80.95	79.55
DSA-HRNet-Inner-4	4	6368	76.65	93.56	84.63	74.12	81.10	79.63
DSA-HRNet-Inner-8	8	6296	76.75	93.57	84.63	74.20	81.24	79.53

In total, results of the four different inner channels are presented. It can be seen that, when the “Inner Channel” decreases, the parameters of the block decrease. Although the mAP drops a bit when the inner channel becomes two or four, when the inner channel becomes eight, it outperforms the original DSA-HRNet. The best AP75 appears when the inner channel is two; meanwhile, the one with inner channel of four produces the best AR. Due to the fact that this inner-channel number determines the quality of feature mapping, conducting hyperparameter tuning enhances the model’s performance.

Conclusion and Future Work

This paper focuses on improving the existing convolutional DNN architecture for solving the HPE problem. It identifies key issues with current methods and proposes the DSA-HRNet, a novel architecture integrating the DSA mechanism into the HRNet that allows for parallel

https://doi.org/10.18080/jtde.v12n3.984

convolutions to maintain HR representations and capture contextual information effectively. Experimental results on the COCO Val 2017 dataset show that DSA-HRNet outperforms the original HRNet, achieving an increase of 2.3%, 3% and 2.7% in terms of mAP, AP50, and AP75, respectively. Moreover, it also compares favourably with other state-of-the-art methods in HPE. Additional experiments confirm the effectiveness of DSA-HRNet, emphasising the importance of hyperparameter tuning for optimal performance.

Some future directions to extend the current work include: firstly, applying the DSA-HRNet to the multi-person pose estimation problem to detect multiple human instances simultaneously alongside single-person pose estimation; secondly, enhancing the DSA module itself through hyperparameter tuning or modifying its inner structure for improved performance. Lastly, while the DSA-HRNet demonstrates excellence on the COCO Val 2017, exploring additional application scenarios, such as applying it to specific pose estimation tasks, could be a direction of future research.

References

Artacho, B., & Savakis, A. (2020). Unipose: Unified human pose estimation in single images and videos. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 7035–7044. https://doi.org/10.48550/arXiv.2001.08095

Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473. https://doi.org/10.48550/arXiv.1409.0473

Cao, Z., Simon, T., Wei, S. E., & Sheikh, Y. (2017). Realtime multi-person 2D pose estimation using part affinity fields. In Proceedings of the IEEE conference on computer vision and pattern recognition, 7291–7299. https://doi.org/10.48550/arXiv.1611.08050

Carreira, J., Agrawal, P., Fragkiadaki, K., & Malik, J. (2016). Human pose estimation with iterative error feedback. In Proceedings of the IEEE conference on computer vision and pattern recognition, 4733–4742. https://doi.org/10.1109/CVPR.2016.512

Chen, Y., Wang, Z., Peng, Y., Zhang, Z., Yu, G., & Sun, J. (2018). Cascaded pyramid network for multi-person pose estimation. In Proceedings of the IEEE conference on computer vision and pattern recognition, 7103–7112. https://doi.org/10.48550/arXiv.1711.07319

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., & Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929. https://doi.org/10.48550/arXiv.2010.11929

Fu, J., Liu, J., Tian, H., Li, Y., Bao, Y., Fang, Z., & Lu, H. (2019). Dual attention network for scene segmentation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 3146–3154. https://doi.org/10.48550/arXiv.1809.02983

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. In Proceedings of the IEEE international conference on computer vision, 2961–2969. https://doi.org/10.1109/ICCV.2017.322

Li, Y., Wang, C., Cao, Y., Liu, B., Luo, Y., & Zhang, H. (2020). A-HRNet: Attention based high resolution network for human pose estimation. In 2020 Second International Conference on Transdisciplinary AI (TransAI), 75–79. IEEE. https://doi.org/10.1109/TransAI49837.2020.00016

Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., & Zitnick, C. L. (2014). Microsoft COCO: Common objects in context. In Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13, 740–755. https://doi.org/10.48550/arXiv.1405.0312

Luo, Y., Ren, J., Wang, Z., Sun, W., Pan, J., Liu, J., & Lin, L. (2018). LSTM pose machines. In Proceedings of the IEEE conference on computer vision and pattern recognition, 5207–5215. https://doi.org/10.48550/arXiv.1712.06316

Papandreou, G., Zhu, T., Kanazawa, N., Toshev, A., Tompson, J., Bregler, C., & Murphy, K. (2017). Towards accurate multi-person pose estimation in the wild. In Proceedings of the IEEE conference on computer vision and pattern recognition, 4903–4911. https://doi.org/10.1109/CVPR.2017.395

Sun, J., Jiang, J., & Liu, Y. (2020). An introductory survey on attention mechanisms in computer vision problems. In 2020 6th International Conference on Big Data and Information Analytics (BigDIA), 295–300. IEEE. https://doi.org/10.1109/BigDIA51454.2020.00054

Sun, K., Xiao, B., Liu, D., & Wang, J. (2019). Deep high-resolution representation learning for human pose estimation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 5693–5703. https://doi.org/10.48550/arXiv.1902.09212

Tan, V. W. S., Ooi, W. X., Chan, Y. F., Connie, T., & Goh, M. K. O. (2024). Vision-Based Gait Analysis for Neurodegenerative Disorders Detection. *Journal of Informatics and Web Engineering, 3*(1), 136–154. https://doi.org/10.33093/jiwe.2024.3.1.9

Ti, Y. F., Connie, T., & Goh, M. K. O. (2023). GenReGait: Gender Recognition using Gait Features. *Journal of Informatics and Web Engineering, 2*(2), 129–140. https://doi.org/10.33093/jiwe.2023.2.2.10

Toshev, A., & Szegedy, C. (2014). DeepPose: Human pose estimation via deep neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, 1653–1660. https://doi.org/10.1109/CVPR.2014.214

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems, 30*, 5998–6008. https://doi.org/10.48550/arXiv.1706.03762

Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018). Deep learning for computer vision: A brief review. *Computational intelligence and neuroscience, 1*, 7068349. https://doi.org/10.1155/2018/7068349

Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., Liu, D., Mu, Y., Tan, M., Wang, X., Liu, W., & Xiao, B. (2020). Deep high-resolution representation learning for visual recognition. *IEEE transactions on pattern analysis and machine intelligence, 43*(10), 3349–3364. https://doi.org/10.1109/TPAMI.2020.2983686

Wang, X., Tong, J., & Wang, R. (2021). Attention refined network for human pose estimation. *Neural Processing Letters, 53*(4), 2853–2872. https://doi.org/10.1007/s11063-021-10523-9

Wu, N., Gao, H., Wang, P., Li, X., & Lv, Z. (2023). High-resolution human pose estimation based on location awareness. In Third International Symposium on Computer Engineering and Intelligent Communications (ISCEIC 2022), Vol. 12462, 129–135. SPIE. https://doi.org/10.1117/12.2660942

Yuan, Y., Fu, R., Huang, L., Lin, W., Zhang, C., Chen, X., & Wang, J. (2021). HRFormer: High-resolution transformer for dense prediction. arXiv preprint arXiv:2110.09408. https://doi.org/10.48550/arXiv.2110.09408

Feature-Level Fusion Multi-Sensor Aggregation Temporal Network for Smartphone-Based Human Activity Recognition

Sarmela Raja Sekaran
Faculty of Information Science and Technology, Multimedia University, Malaysia

Pang Ying Han
Faculty of Information Science and Technology, Multimedia University, Malaysia

Lim Zheng You
Faculty of Information Science and Technology, Multimedia University, Malaysia

Ooi Shih Yin
Faculty of Information Science and Technology, Multimedia University, Malaysia

Abstract: Smartphone-based Human Activity Recognition (HAR) identifies human movements using inertial signals gathered from multiple smartphone sensors. Generally, these signals are stacked as one (data-level fusion) and fed into deep learning algorithms for feature extractions. This research studies feature-level fusion, individually processing inertial signals from each sensor, and proposes a lightweight deep temporal learning model, Feature-Level Fusion Multi-Sensor Aggregation Temporal Network (FLF-MSATN), that performs feature extraction on inertial signals from each sensor separately. The raw signals, segmented into equally sized time windows, are passed into individual Dilated-Pooled Convolutional Heads (DPC Heads) for temporal feature analysis. Each DPC Head has a spatiotemporal block containing dilated causal convolutions and average pooling, to extract underlying patterns. The DPC Heads' outputs are concatenated and passed into a Global Average Pooling layer to generate a condensed confidence map before activity classification. FLF-MSATN is assessed using a subject-independent protocol on a publicly available HAR dataset, UCI HAR, and a self-collected HAR dataset, achieving 96.67% and 82.70% accuracies, respectively. A Data-Level Fusion MSATN is built to compare and verify the model performance attained by the proposed FLF-MSATN. The empirical results show that implementing FLF-MSATN enhances the accuracy by ~3.4% for UCI HAR and ~9.68% for self-collected datasets.

https://doi.org/10.18080/jtde.v12n3.979

Keywords: smartphone, inertial signal, deep learning, multi-sensor, dilated convolution

Introduction

In recent decades, human activity recognition (HAR) has garnered much interest from researchers, as HAR has extensive applications in countless domains, such as ambient assisted living (Wan *et al.*, 2021), gait recognition (Tan *et al.*, 2024; Ti *et al.*, 2023), sports (Stein *et al.*, 2016), and security (Ben Mabrouk & Zagrouba, 2018). Visual-based and sensor-based HAR are the major categories in the HAR domain. Although visual-based HAR is a widely researched domain, this technique has its pitfalls. Visual-based HAR requires high computational resources to process and store two- or three-dimensional input data. Additionally, the performance of visual-based HARs relies on the input quality to a great extent, which is easily disrupted by occlusions, illumination settings and non-human motion in the surroundings (Perez-Gamboa *et al.*, 2021). Moreover, this technique violates user privacy, as the input data contains sensitive user identification information.

Unlike visual-based HAR, sensor-based HAR requires relatively lower computational resources for processing, due to its one-dimensional input data, and its performance is not influenced by background noises, illuminations or occlusions (Perez-Gamboa *et al.*, 2021). Most importantly, sensor-based HAR preserves a user's privacy, as it does not collect user identity information (Yang *et al.*, 2015). Smartphone-based HAR is more prevalent than other sensor-based HAR for several reasons. Nowadays, almost every individual already carries a smartphone, so data collection is much easier and more cost-effective among diverse populations, as it does not require additional sensors. Moreover, data collection is more convenient and less obtrusive for users as they are not required to wear additional sensors or devices. Hence, this research work focuses on developing a smartphone-based HAR system.

Data scientists have developed numerous techniques (e.g., handcrafted feature-based algorithms, deep learning algorithms, or both) to build robust smartphone-based HAR systems. Although the handcrafted feature-based methods produced adequate performance, they have taken a backseat due to their time-consuming processes, including manual feature engineering, ranking, and selections (Bengio, 2013; Minh Dang *et al.*, 2020). The generalisability of handcrafted feature-based methods is low, due to their dependency on prior knowledge. Hence, deep learning algorithms have been extensively implemented, as they can automatically capture salient features from input data for classification. The most popular deep learning algorithms are Long Short-Term Networks (LSTM) (Alawneh *et al.*, 2020; Ullah *et al.*, 2019), Gated Recurrent Units (GRU) (Dua *et al.*, 2021; Mim *et al.*, 2023), Temporal Convolutional Networks (TCN) (Garcia *et al.*, 2019; Raja Sekaran *et al.*, 2023) and

Convolutional Neural Networks (CNN) (Huang *et al.*, 2020; Khan & Ahmad, 2021; Teng *et al.*, 2020).

Related work

Recently, researchers have invented deep learning algorithms to build a high-performing smartphone-based HAR system, due to the abundance of computing power. Huang *et al.* addressed the low interclass variance problem between certain classes and proposed a deep learning model with two CNN networks, TSE-CNN, to improve the classification performance (Huang *et al.*, 2020). Each CNN contains two convolutions and a pooling layer. The first model classifies upstairs, downstairs, and other activities, while the second model takes in the input data classified as other activities and then classifies them into walking, jogging, standing, and sitting. In Zhang et al. (2020), the authors built a multi-head CNN model followed by attention modules to classify raw input signals into respective activity classes. This model extracts features at multiple scales, and the attention modules will select relevant features and remove redundant information.

Aside from CNNs, recurrent networks are most desirable for time-series classification as these networks are good at extracting temporal dependencies from input signals, which is crucial in human motion analysis. Ullah *et al.* designed a multi-layer LSTM model to perform HAR without manual feature preprocessing (Ullah *et al.*, 2019). The authors placed a single neural network layer before the stacked LSTM fitted with an L2 regulariser to capture salient patterns from the input. Alawneh *et al.* presented two recurrent models, unidirectional LSTM and bidirectional LSTM, to identify which model performs well in the HAR task (Alawneh *et al.*, 2020). The authors found that bidirectional LSTM achieved higher accuracy than unidirectional LSTM on UniMiB SHAR and WISDM datasets, as this architecture utilises previous and subsequent information of the input sequence during classification.

Many researchers have begun to combine CNN with recurrent networks to leverage benefits from both models. For instance, Dua *et al.* designed a multi-input architecture amalgamating a CNN model and a GRU model and tested their architecture on WISDM, PAMAP2 and UCI HAR (Dua *et al.*, 2021). The authors built three heads, each comprising convolutional layers, max pooling and GRU layers. The authors found that multi-input CNN-GRU outperforms existing models in all three datasets. Likewise, Xia *et al.* developed a hybrid network combining an LSTM model with convolutional layers (Xia *et al.*, 2020). The authors stated that the hybrid network can efficiently learn spatiotemporal dynamics from the input sequences and make more accurate predictions.

Another new architecture using one-dimensional convolutional layers, TCN, has been designed for time series classification (i.e., weather forecasting, video recognition, speech recognition, etc.). TCNs can capture longer temporal features with a relatively low computational cost compared to recurrent networks (Bai *et al.*, 2018). Nair *et al*. adopted two TCN architectures from the work of Lea et al. (2016) to train and test them with the HAR dataset (Nair *et al.*, 2018). These TCNs could learn temporal patterns from raw signals and achieved higher accuracy than most existing models on UCI HAR. Raja Sekaran *et al.* developed an Inception-inspired TCN model to perform classification without complex preprocessing (Raja Sekaran *et al.*, 2022). This model can extract long temporal patterns and feature maps at different scales due to the varying convolutional kernel sizes.

Besides model architecture, input data is a significant determining factor for achieving high-accuracy classifiers. Nowadays, almost every smartphone includes accelerometers, gyroscopes, and magnetometers that can be utilised for HAR tasks. There are two types of HAR datasets: single-sensor (i.e., WISDM and UniMiB SHAR) and multi-sensor (i.e., UCI HAR). When the HAR systems are built on single-sensor data, they are bound to suffer from limited or loss of crucial details caused by limited spatial coverage, sensor deprivation, uncertainty and imprecision (Gravina *et al.*, 2017). This may lead to poor model performance.

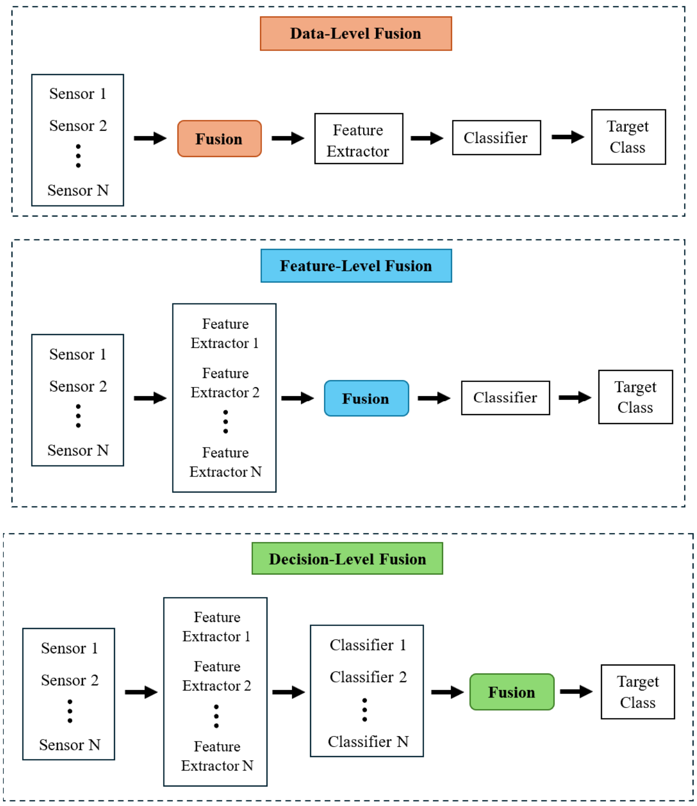


Figure 1. Types of data fusions

In contrast, multi-sensor data provides richer information that a single sensor may have missed (Aguileta *et al.*, 2019). However, the main challenge in developing a multi-sensor HAR system is deciding how and when to combine the inertial signals from different sensors for feature extraction. Figure 1 shows the three ways to combine multi-sensor signals: data-level, feature-level, and decision-level fusion. The multi-sensor signals are concatenated during data preprocessing before passing into the feature extractor in data-level fusion. As for feature-level fusion, multi-sensor signals are fed into individual feature extractors, and the extracted representations are concatenated to be classified. Similarly, multi-sensor signals are preprocessed and undergo feature extraction individually in decision-level fusion. However, unlike previous fusions, the extracted features are classified using individual classifiers, and then the individual decisions are combined to make the final prediction.

Motivations and contributions

Though various approaches have been devised to design a smartphone-based multi-sensor HAR system, there is still much room for improvement. The following are several hurdles in achieving a high-performing HAR model:

- Combining multi-sensor inertial signals at an early stage (data-level fusion) may lead to sensor-specific information loss, where each sensor may possess unique characteristics that allow the classifier to easily differentiate between different activities, as illustrated in Figure 2. This fusion also increases the input dimensionality, which raises the complexity.
- Modelling longer-term dependency remains a challenge in the HAR domain. For example, handcrafted feature-based approaches depend on manual feature engineering/selection based on prior knowledge, which may cause the model to miss the underlying temporal information in the input sequences. They are also unable to generalise to new unseen samples. Moreover, traditional CNNs are also suboptimal in long temporal pattern extraction, as they are designed for spatial feature extraction (Asim *et al.*, 2017).
- Since deep learning algorithms have multi-layered architectures and contain a large number of parameters, they usually require a high computational resource, resulting in higher computational costs and the possibility of model overfitting. For example, LSTMs implement multiple gate operations at each layer, so processing and storing the intermediate results require high computational resources (Bai *et al.*, 2018). Though TCNs are relatively less computationally demanding, in order to achieve desirable performance, they require multiple one-dimensional convolutions involving plenty of huge convolutional kernels, which results in deeper and denser networks (Garcia *et al.*, 2019; Nair *et al.*, 2018; Raja Sekaran *et al.*, 2022).

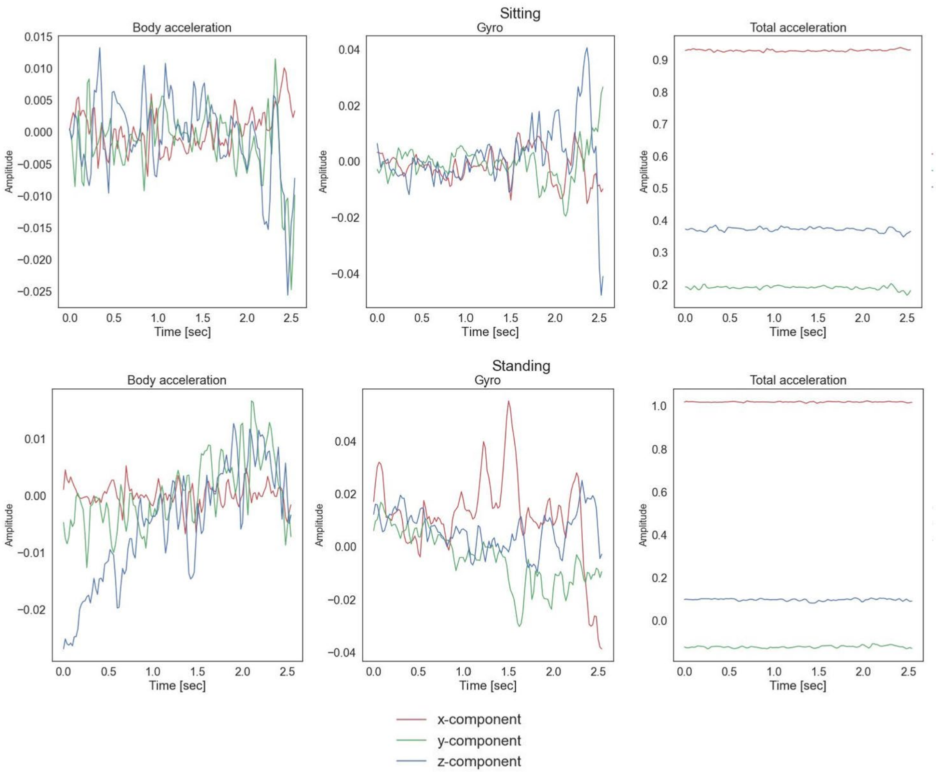


Figure 2. Signal samples of sitting and standing classes

Given the previously mentioned challenges, we propose a Feature-Level Fusion Multi-Sensor Aggregation Temporal Network (FLF-MSATN), a lightweight multi-sensor deep learning architecture, to classify human actions, as illustrated in Figure 3. The contributions of this research are:

- The proposed FLF-MSATN extracts salient features from the multi-sensor signal individually using multiple Dilated-Pooled Convolutional Heads (DPC Heads) before fusing them (i.e., feature-level fusion) to gain more significant insight into the sensor-specific information. Each DPC Head receives sensor-specific input signals and processes them to capture distinctive implicit patterns. Then, the outputs of the DPC Heads are concatenated to ensure that various features from each sensor-specific signal are combined without altering any captured features before final classification.
- The spatiotemporal block in the DPC Head comprises sequentially organised one-dimensional dilated causal convolutional layers with a larger convolutional kernel and increasing dilation rate to capture and retain long temporal dependencies. Capturing longer temporal dependent features allows the proposed FLF-MSATN to gain more information regarding the characteristics of each activity. Additionally, causality is introduced to preserve the input sequence's ordering.
- The proposed FLF-MSATN is a lightweight deep learning-based HAR model. Implementing dilations in DPC Heads drastically reduces the overall learnable parameters of the proposed FLF-MSATN, lowering the computational complexity and

costs. Furthermore, applying the Global Average Pooling (GAP) layer minimises the total parameters and model overfitting tendency.

- The proposed FLF-MSATN is validated on a popular HAR dataset, UCI HAR, for benchmarking purposes and a self-collected dataset using a user-independent protocol. In this protocol, the training and test set do not share samples from the same user to replicate real-world scenarios.

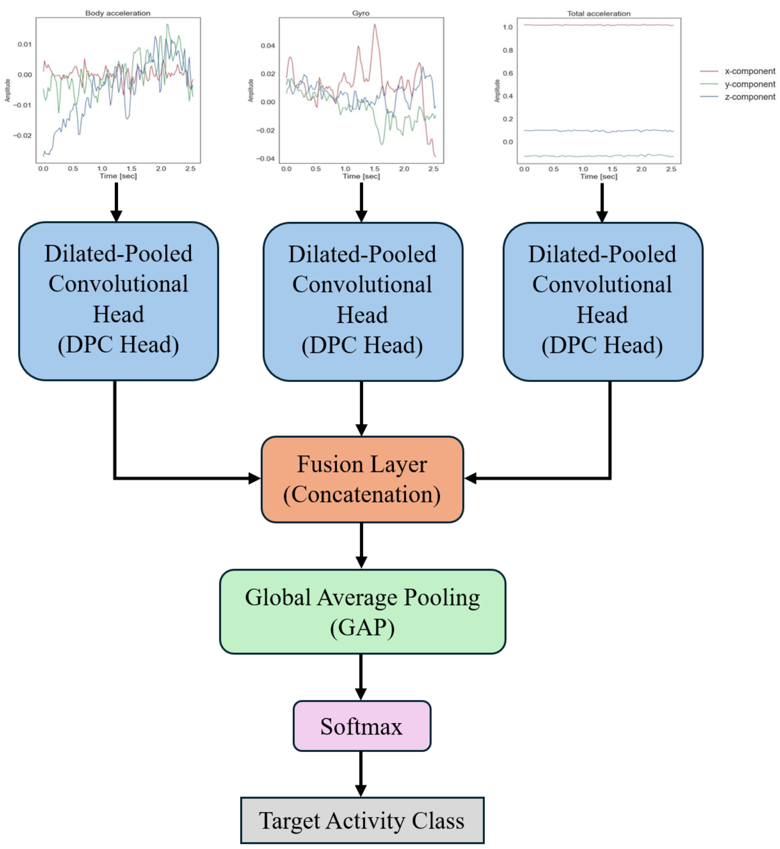


Figure 3. The overall architecture of the proposed FLF-MSATN

Proposed Architecture

The proposed FLF-MSATN comprises DPC Heads, one fusion layer, one GAP layer and an output layer with softmax activation. The components are organised in a specific order to allow the proposed FLF-MSATN to automatically learn salient sensor-specific features, improving the classification performance. The following subsections describe the proposed architecture in depth.

Data segmentation

As previously mentioned, two datasets (e.g., UCI HAR and self-collected datasets) validate the proposed FLF-MSATN. The details of both datasets are presented in the subsection "Dataset description" under section "Experimental Setups". Both datasets utilise smartphone sensors, like accelerometer and gyroscope, to gather inertial signals. Each dataset has three triaxial inertial data (i.e., linear and total acceleration and gyroscope readings). Since triaxial input signals cannot be passed directly into the feature extractor, these signals are segmented using the sliding window technique into equally sized time windows, where each window has 128 data points. Then, these time windows are passed into the sensor-specific DPC Head for spatiotemporal feature extraction.

Feature-level fusion

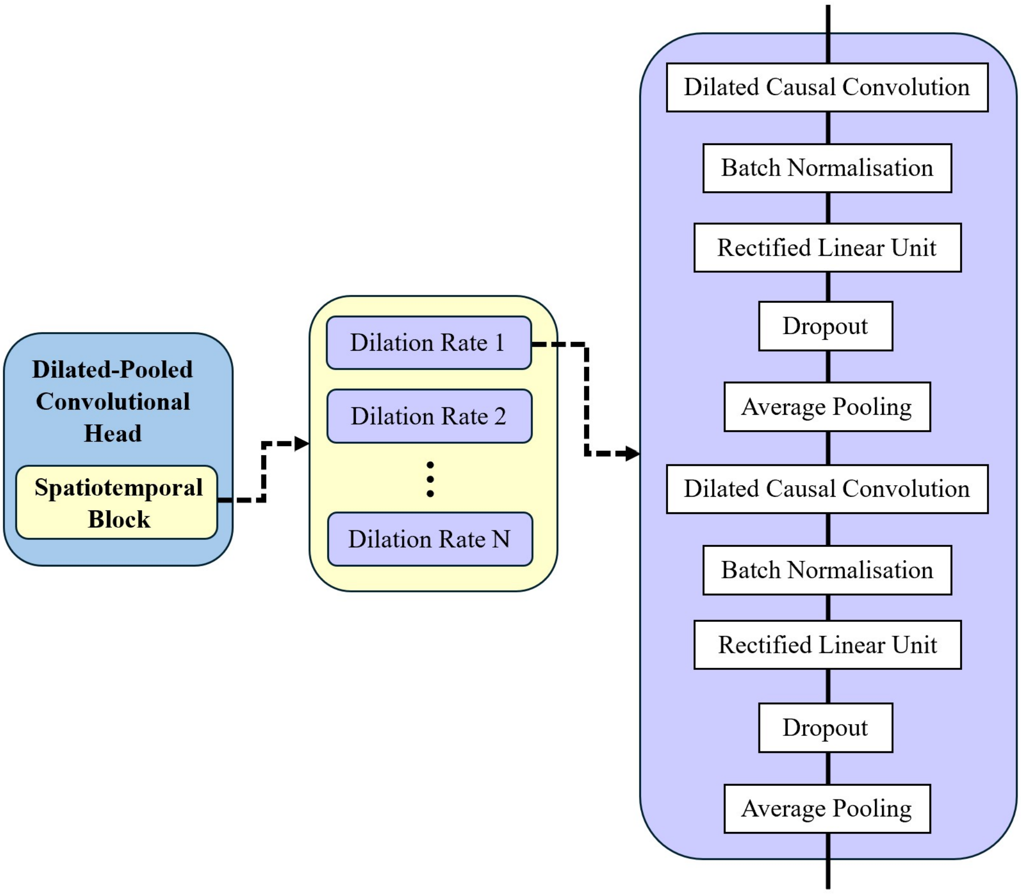


Figure 4. DPC Head architecture

The most significant component of the proposed FLF-MSATN is the sensor-specific DPC Heads, organised in parallel. Figure 4 shows the complete architecture of the DPC Head. DPC Heads comprise a spatiotemporal block with two one-dimensional dilated causal convolutional layers, where a batch normalisation, activation function, dropout, and average pooling layer follow each. The backbone of the spatiotemporal block is the one-dimensional dilated causal convolutional layers. When longer convolutional kernels are implemented, the model can capture more extended temporal information from the input sequence. Nevertheless, this setting drastically raises the model's trainable parameters, increasing the computational load. Thus, integrating dilation into the convolutions aids the proposed architecture in extending its receptive field by adding zeros between the kernel values, as illustrated in Figure 5, to extract longer information during the convolution process without increasing the total parameters. Additionally, causal padding is incorporated into the dilated convolutions to preserve the input signal's ordering.

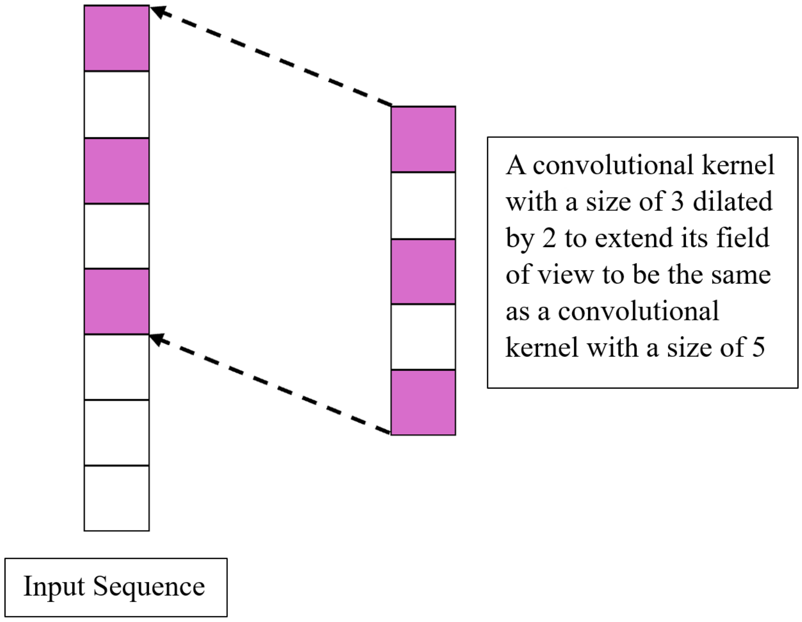


Figure 5. Dilated convolution operation

Batch normalisation is employed in the proposed FLF-MSATN to stabilise model training by reducing the model's internal covariate shift. During training, the model is fed with batches of inputs. In each batch, mean and variance are computed in the batch normalisation layer to normalise the output vectors. However, two learnable parameters in that layer are used to scale and shift the normalised output vectors, ensuring that the mean and variance of each batch stay constant (Ioffe & Szegedy, 2015).

Moreover, the ReLU (Rectified Linear Unit) function, a non-linear activation, is implemented to convert negative values into true zeroes. Zero values allow the network to maximise its sparsity, as the model then becomes less complex, intrinsically speeding up the training process (Rasamoelina *et al.*, 2020). ReLU activation is low in computation (Glorot *et al.*, 2011) and reduces the gradient vanishing problem (Mercioni & Holban, 2020). The formula for this activation can be written as follows:

$$f(x) = \max(0, x) \tag{1}$$

where x is the input to ReLU activation. The output of the ReLU activation ranges from zero to infinite. For instance, if the input is positive, this activation outputs the same fed value, whereas, if the input is negative, then this activation returns zero.

Another issue in deep learning models is model overfitting. The classifiers are considered overfitted if fitted too closely to the training examples and cannot generalise properly, leading to poor performance on unseen test samples. Generally, overfitting is caused by either the classifier being too complex or the size of the training set being too low (Xu *et al.*, 2019). Hence, the proposed FLF-MSATN incorporates regularisation to avoid this phenomenon. Each dilated causal convolution is fitted with a dropout layer (rate of 0.05), as shown in Figure 4. Randomly ignoring the hidden layer's neurons according to the dropout rate introduces sparsity and simplifies the overall network.

In addition, the DPC Head integrates average pooling to aid input dimensionality reduction, resulting in fewer learnable parameters (Zafar *et al.*, 2022). The average pooling accepts convolved feature maps and applies an average filter over the non-overlapping regions to generate the averaged feature maps, as shown in Figure 6. This pooling is adopted into the proposed FLF-MSATN because it considers every value of the feature maps to generate an output vector, which prevents information loss.

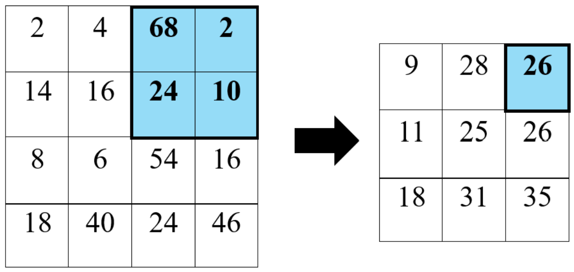


Figure 6. Average pooling operation

Each DPC Head is a spatiotemporal feature extractor for each sensor-specific segmented data. Then, the extracted output feature maps of the DPC Heads will be passed into the fusion layer. All the output feature maps are concatenated, instead of combined using arithmetical functions, like addition, subtraction, or multiplication, to avoid the loss of underlying details before the classification stage.

Classification

Conventionally, a couple of fully-connected dense layers process the extracted features prior to activity classification, but implementing these leads to several pitfalls that negatively impact the model performance. The fully-connected dense layers substantially increase the learnable parameters, escalating the model's complexity. Inevitably, a complex model requires more resources during training, which makes it more susceptible to model overfitting. Moreover, fully-connected dense layers also demand parameter optimisation to achieve better performance.

Since this study focuses on designing lightweight deep learning, the fully-connected dense layers are replaced with a GAP layer. GAP layers do not raise the model's trainable parameters, as these layers generate a condensed one-dimensional feature map from multi-dimensional feature maps. For example, the GAP layer accepts 96 channels of one-by-two input vector and

https://doi.org/10.18080/jtde.v12n3.979

converts them into a one-by-one output vector with 96 channels, removing many learnable parameters. Furthermore, the GAP layer does not require hyperparameter optimisation (Xia *et al.*, 2020).

Finally, the last dense layer is equipped with units that are the same as the total activity classes and softmax activation. This layer accepts the condensed feature maps for activity classification. The softmax function generates class probabilities that fluctuate between zero and one, where the activity's probability closest to one is identified as the target activity. Softmax function, $\alpha()$, can be defined as follows:

$$\alpha(\vec{x}) = \frac{e^{x_i}}{\sum_{j=1}^{N} e^{x_j}} \tag{2}$$

where $\vec{x}$ is the input vector, e^{x_i} is the standard exponential function of i^{th} element of the input vector and N is the total number of activities. The model training implements the categorical cross-entropy, a multi-class loss function, allowing the proposed model to learn from its misclassification errors by adjusting the weights of layers and making more accurate classification.

Experimental Setups

The following subsections describe the databases used for evaluation and the parameter configuration of the proposed architecture.

Dataset description

UCI HAR dataset is contributed to the UC Irvine Machine Learning Repository by Anguita *et al.* (2013) as a benchmarking smartphone-based HAR dataset and can be accessed at https://archive.ics.uci.edu/dataset/240/human+activity+recognition+using+smartphones. This dataset comprises 10299 samples collected from thirty users aged 19 to 48. These samples are triaxial signals collected using the Samsung Galaxy S II smartphone, embedded with an accelerometer and gyroscope sensor. The users performed the necessary activities with the smartphone placed on their waist. The authors implemented a subject-independent procedure and randomly segregated the whole dataset into training and test sets, where samples from twenty-one users are placed in the training set and the remaining ones in the test set.

Besides UCI HAR, we developed a data collection application to gather triaxial inertial signals using the Samsung S8 Plus smartphone. The accelerometer and gyroscope are the two sensors selected from the smartphone to perform the data collection. A total of 12312 samples were collected from thirteen volunteers aged above 60. During the data collection, the volunteers had to place the smartphone in their left pants pocket and perform each activity for two

https://doi.org/10.18080/jtde.v12n3.979

minutes. The activity classes are walking, climbing stairs, sitting, standing, brisk walking, and laying. We also obeyed the subject-independent procedure. All the samples from the randomly selected nine users are placed on the training set, and the remaining ones on the test set. The full details of UCI HAR and self-collected databases are presented in Table 1.

Table 1. Description of UCI HAR and the self-collected databases

Characteristics	**Datasets**	
	UCI HAR	**Self-collected Database**
Sensors Involved	Accelerometer and gyroscope	Accelerometer and gyroscope
Sampling Rate	50Hz	20 Hz
Number of Volunteers	30	13
Activity Classes	Walking, walking upstairs, sitting, standing, and laying	Walking, climbing stairs, sitting, standing, brisk walking and laying
Input Segment Size	128	128
Channel Size	9	9
Training-Test Split	70:30 (21 training subjects: 9 test subjects)	70:30 (9 training subjects: 4 test subjects)
Validation Split	10% of the training data	20% of the training data
Validation Procedures	Subject-independent	Subject-independent

Model configuration

The proposed FLF-MSATN is designed using open-source machine learning libraries, such as TensorFlow 2.4.1 and Keras 2.4.0, where the codes are written in Python 3.8. The model evaluations are carried out on a desktop running on the 64-bit operating system of Windows 10 with Intel® Core™ i9-12900K CPU with 2.20 GHz, 32GB RAM, NVIDIA GeForce RTX 3080Ti and 12GB memory. A Data-Level Fusion Multi-Sensor Aggregation Temporal Network (DLF-MSATN), as illustrated in Figure 7, is also built to study whether feature-level fusion is better than data-level fusion. The model configurations of the proposed FLF-MSATN and DLF-MSATN for both datasets are presented in Table 2. During the ablation studies, the optimal hyperparameter values of the proposed FLF-MSATN are determined (discussed in the subsection "Architectural ablation study").

The number of training epochs is set to 100 as the model performance (i.e., accuracy and loss) plateaued after 100 epochs. With a 0.001 initial learning rate, Adam Optimiser is utilised to train the models. Unlike conventional training, a dynamic learning rate strategy is implemented using a pre-built TensorFlow function, Reduce Learning Rate on Plateau, to speed up model convergence. The idea behind this function is that the learning rate will be reduced by a specified factor when validation loss or accuracy stops improving over a fixed number of training epochs, allowing the model to learn continuously. Besides the dynamic learning rate, model training also includes ModelCheckpoint functionality, which saves the model with optimal performance. This function is able to decide the optimality of a model based on the model's validation loss or accuracy. For instance, if ModelCheckpoint mode is set to validation loss, this function will store the model whenever the validation loss drops during the training. Storing optimal models allows one to select a model for testing before it overfits.

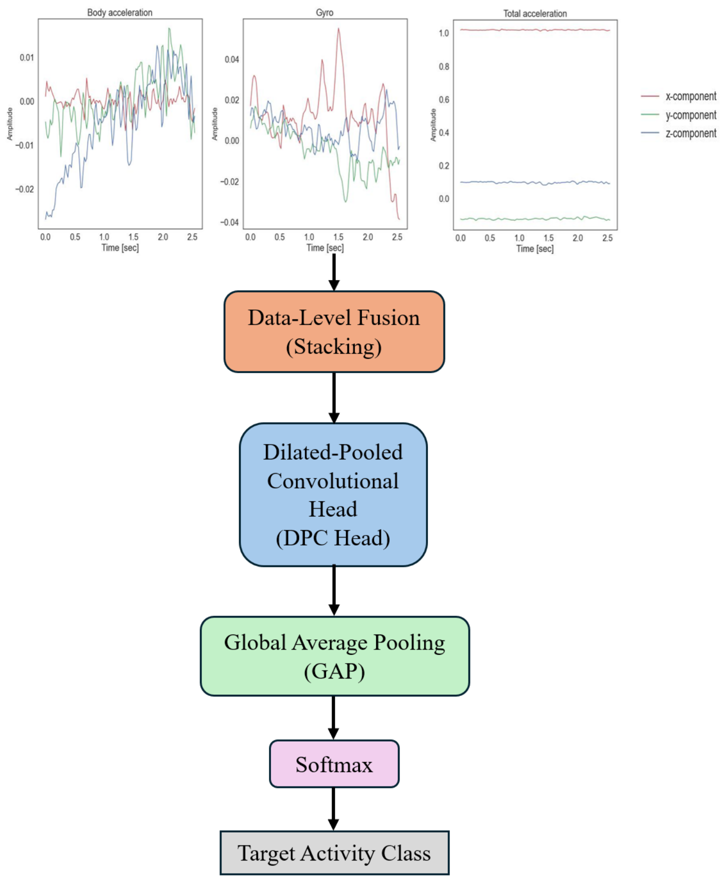


Figure 7. DLF-MSATN architecture

Table 2. Parameter configuration of the proposed FLF-MSATN and DLF-MSATN for UCI HAR and self-collected databases

Parameter	**FLF-MSATN**		**DLF-MSATN**	
	UCI HAR	**Self-collected**	**UCI HAR**	**Self-collected**
Input Shape	(128,3,3)	(128,3,3)	(128,9)	(128,9)
Batch Size	32	32	32	32
Total Training Epochs	100	100	100	100
Initial Learning Rate	0.001	0.001	0.001	0.001
Number of spatiotemporal blocks in DPC Head	1	1	1	1
Number of Kernels	32	32	32	32
Kernel Size	3	8	3	8
Stride	1	1	1	1
Dilation Rate	1,2,4	1,2,4	1,2,4	1,2,4
Pooling Size	2	2	2	2
Dropout Rate	0.05	0.05	0.05	0.05
Optimiser	Adam	Adam	Adam	Adam

Parameter	FLF-MSATN		DLF-MSATN	
	UCI HAR	Self-collected	UCI HAR	Self-collected
Dynamic Learning Rate Settings	Mode: Validation Loss Patience: 3 Factor: 0.5 Minimum Learning Rate: 0.0001	Mode: Validation Loss Patience: 3 Factor: 0.5 Minimum Learning Rate: 0.0001	Mode: Validation Loss Patience: 3 Factor: 0.5 Minimum Learning Rate: 0.0001	Mode: Validation Loss Patience: 3 Factor: 0.5 Minimum Learning Rate: 0.0001
Loss function	Categorical Cross-entropy	Categorical Cross-entropy	Categorical Cross-entropy	Categorical Cross-entropy

Results and Discussions

The subsequent subsections describe the architectural ablation study, the proposed FLF-MSATN's performance evaluation, and the results analysis.

Architectural ablation study

Hyperparameter optimisation is essential in designing a deep learning-based HAR model. Determining optimal values, parameters, or architectures will enhance the overall model performance. Acquiring a balanced trade-off between achieving a high model performance (i.e., classification accuracy) and computational efficiency (i.e., low trainable parameters) is the main motive of this ablation study. Three experiments are conducted on the proposed FLF-MSATN using the self-collected database, including kernel sizes, feature-level fusion type, and implementation of the GAP layer.

Table 3. Model performance with different convolutional kernel size settings

Kernel Size	Number of Parameters (Million)	F1 Score	Accuracy (%)
4	0.065	0.8043	81.96
8	0.128	0.8102	82.70
12	0.191	0.7968	82.10
16	0.253	0.7816	80.27

Extracting longer temporal features is essential in HAR for analysing and identifying unique and discriminative patterns from inertial signals that can distinguish between activities easily (Ismail Fawaz *et al.*, 2020). A conventional way to capture longer temporal patterns is kernel extension. Nevertheless, integrating longer convolutional kernels results in more model parameters, increasing the model's vulnerability to overfitting (Ismail Fawaz *et al.*, 2020). Hence, an experiment is conducted to determine an appropriate kernel size for the proposed FLF-MSATN. Table 3 records the experimental results for different kernel size settings. The overall trainable parameters increased as the kernel increased. The performance of the proposed FLF-MSATN improved by ~1% when the kernel size was increased from four to eight. However, the performance declines when the kernel size is set to twelve or sixteen. The proposed FLF-MSATN had the best classification performance (i.e., 0.8102 F1 Score and 82.70% accuracy) with the lowest total parameters (i.e., 0.128 million parameters) when the kernel size is eight compared to other sizes. The model with a kernel size of eight achieved a balanced trade-off between classification performance and the model complexity.

Since the proposed FLF-MSATN is a multi-sensor HAR, the proposed model is tested with different feature-level fusion techniques. The outputs of the sensor-specific DPC Heads are fed into the fusion layer, where they will be combined for activity classification. Those features either undergo arithmetic addition or concatenation at the fusion layer. In the addition technique, the outputs of the DPC Heads are summed together and encoded as a new feature representation, whereas, in the concatenation method, the outputs of the DPC Heads are appended to form a new feature representation without modifying the existing feature maps. Table 4 presents the proposed model's performance under different settings. The proposed FLF-MSATN with concatenation as the fusion layer outperformed the model with the arithmetic addition layer by ~3.4%. This result may be due to the ability of the concatenation layer to combine the feature maps without manipulating the existing feature maps, preserving the underlying temporal patterns of the sensor-specific feature maps.

Table 4. Model performance with different feature-level fusion

Feature-level Fusion Type	F1 Score	Accuracy (%)
Addition	0.7731	79.26
Concatenation	0.8102	82.70

The proposed FLF-MSATN implements the GAP layer after the fusion layer, as illustrated in Figure 3. The fused feature maps are condensed into compact confidence maps in the GAP layer. An experiment is designed to identify whether the GAP layer has an advantage over fully-connected dense layers regarding classification performance and number of parameters. Table 5 records the experimental results. Indeed, the model with the GAP layer acquired higher accuracy and a lower overall parameter count. Replacing a fully-connected dense layer with a GAP layer removed a chunk of model parameters (i.e., ~ 40000 trainable parameters) from the proposed FLF-MSATN. Additionally, there is no need for parameter optimisation for the GAP layer.

Table 5. The impact of the GAP layer on the model performance

Implementation of GAP Layer	Number of Parameters (Million)	F1 Score	Accuracy (%)
Yes	0.128	0.8102	82.70
No (Fully-connected dense layer)	0.162	0.7570	79.12

Model performance evaluation and comparison

The following subsections will discuss the performance analysis and comparison between the proposed FLF-MSATN and other HAR architectures.

UCI HAR

This work has designed a deep learning model that treats sensor-specific inertial signals independently. Hence, to verify whether the feature-level fusion architecture improves the model performance, DLF-MSATN, a model that accepts input signals from different sensors that already have been fused before feature extraction, as mentioned in subsection "Model configuration", is trained and tested using the same procedures as the proposed FLF-MSATN. The proposed FLF-MSATN achieved higher classification accuracy than the DLF-MSATN by ~3.4%, as shown in Table 6. Notably, the proposed FLF-MSATN makes better predictions, as it retains the unique characteristics of the sensor-specific inertial signals, whereas DLF-MSATN may be susceptible to information loss. To further understand the feature extraction process of these models, the feature representations of both models from the same layer are illustrated in Figure 8, with the colour generated based on the value of the feature representations. The proposed FLF-MSATN has three times more information than DLF-MSATN, indicating that the feature-level fusion model provides richer information for classification.

Table 6. Performance comparison of the proposed FLF-MSATN with existing HAR architectures on UCI HAR

Methods	Accuracy (%)	The Number of Learnable Parameters (Million)
Lego Filter CNN (Tang *et al.*, 2021)	96.90*	1.300*
RestHAR (Shome, 2021)	95.20*	-
Attention induced multi-head CNN (Khan & Ahmad, 2021)	95.38*	1.510*
Stacked LSTM (Ullah *et al.*, 2019)	93.13*	-
Encoder-Decoder TCN (Nair *et al.*, 2018)	94.60*	0.160*
Dilated TCN (Nair *et al.*, 2018)	93.80*	0.150*
Light-MHTCN (Raja Sekaran *et al.*, 2023)	96.47*	0.210*
ISPLInception (Ronald *et al.*, 2021)	95.09*	1.330*
LSTM-CNN (Xia *et al.*, 2020)	95.78*	0.050*
CNN-BiLSTM (Challa *et al.*, 2022)	96.37*	0.631*
GRU-INC (Mim *et al.*, 2023)	96.20*	0.670*
DLF-MSATN	93.25	0.017
Proposed FLF-MSATN	**96.67**	**0.049**

*Results extracted from the respective articles

Table 6 records performance comparison (i.e., accuracy and overall model parameters) between the proposed FLF-MSATN and other existing models. The performance of the proposed FLF-MSATN exceeds most deep learning algorithms, except CNN with the Lego Filters model, by achieving an accuracy of 96.67% with only 0.049 million parameters.

https://doi.org/10.18080/jtde.v12n3.979

Although GRU-INC, CNN-BiLSTM and Light-MHTCN models only had slightly poorer results, the proposed model is still considered superior to them, as these models require a large chunk of learnable parameters to achieve an adequate model performance (Table 6). LSTM-CNN model had a similar number of learnable parameters as the proposed model, but the proposed FLF-MSATN acquired better accuracy than LSTM-CNN model by ~1.5%. Lego Filter CNN architecture's performance is marginally higher than the proposed FLF-MSATN with a minute difference of ~0.2%. However, the proposed architecture's performance is still on par with Lego Filter CNN model, as this architecture contains 1.3 million parameters, approximately twenty-six times the overall model parameters of the proposed FLF-MSATN.

Figure 8. Feature representation of the proposed FLF-MSATN and DLF-MSATN on UCI HAR

Self-collected dataset

As with the UCI HAR dataset, DLF-MSATN is built with the parameter settings provided in Table 2 and evaluated on the self-collected database. Inevitably, the proposed FLF-MSATN outperformed DLF-MSATN by approximately 9.6%, as the feature-level fusion model preserves the uniqueness of the inertial signals from each sensor.

Furthermore, the proposed FLF-MSATN's performance is compared with the popular HAR algorithms, as presented in Table 7. Several HAR architectures, including traditional machine and deep learning algorithms, are chosen and validated on the self-collected database based on default hyperparameter configurations. From the empirical results, the proposed FLF-MSATN's performance is superior to all the tested models. The traditional machine learning models like SVM variants, K-Nearest Neighbour (K-NN), Decision Tree (DT) and Random Forest classifier (RF) cannot capture implicit spatiotemporal patterns from the input inertial signals that discriminate between the activity classes, resulting in poor classification performance. The CNN model had the lowest accuracy among the tested models due to its lack of capturing temporal patterns. Bi-LSTM and CNN-LSTM achieved the highest accuracy compared to other tested deep learning models. Recurrent models, especially LSTM networks,

https://doi.org/10.18080/jtde.v12n3.979

are good at extracting temporal dependencies from the inertial signals. However, the proposed FLF-MSATN achieved ~5% accuracy more than these models with lower model parameters, as the model captures long underlying temporal patterns from sensor-specific motion signals individually and concatenates them without manipulating the existing features, providing richer insight into the characteristics of each activity for classification.

Table 7. Performance comparison of the proposed FLF-MSATN with existing HAR architectures on the self-collected database

Method	Accuracy (%)	The Number of Learnable Parameters (Million)
Linear SVM	55.36	-
Polynomial SVM	56.68	-
Nu-SVM	67.38	-
K-NN (K=3)	67.30	-
Decision Tree	66.58	-
RF	70.28	-
CNN	63.96	0.226
LSTM	75.15	0.084
Stacked LSTM	71.87	0.263
Bi-LSTM	77.50	0.168
Stacked GRU	65.82	1.072
CNN-LSTM	77.12	0.315
ISPLInception (Ronald *et al.*, 2021)	74.49	0.218
TCN (Lea *et al.*, 2016)	68.14	1.734
TCN (Bai *et al.*, 2018)	71.76	0.283
DLF-MSATN	72.28	0.044
Proposed FLF-MSATN	**82.70**	**0.128**

Conclusions

We put forward a lightweight feature-level fusion deep learning architecture, FLF-MSATN, for smartphone-based HAR. Implementing feature-level fusion aims to preserve unique temporal patterns from multi-sensor signals for feature extraction, allowing the proposed FLF-MSATN to gain deeper insight into motion analysis. The proposed FLF-MSATN is a multi-head network containing multiple DPC Heads, where each DPC Head with a spatiotemporal block receives sensor-specific signals for spatiotemporal feature extraction. A spatiotemporal block contains multiple dilated causal convolutions and average pooling. The convolutional layers implement dilations to expand each kernel's field of view during convolutional operation while maintaining low trainable parameters, improving the information gain and simplifying the model. Each dilated causal convolution is followed by batch normalisation, ReLU activation and dropout regularisation. The batch normalisation maintains the mean and variance of each batch to minimise internal covariate shifts. ReLU activation is responsible for introducing non-linearity into the proposed FLF-MSATN, which enhances model performance. The reason for implementing dropout regularisation is to maximise sparsity in the proposed network, making our model less vulnerable to overfitting. Besides that, the proposed FLF-MSATN incorporates average pooling to reduce the dimensionality of the inputs and remove noise, improving the model performance. Then, the outputs of the DPC Heads are concatenated before passing into the GAP layer, which is also responsible for decreasing the number of model parameters by condensing the fused features to form a compact one-dimensional feature representation.

The proposed FLF-MSATN's efficacy is validated on UCI HAR and self-collected databases using the user-independent protocol. As a result, the proposed FLF-MSATN showed promising classification performance across both datasets compared to other HAR architectures. The proposed FLF-MSATN achieved 96.67% accuracy with 0.049 million parameters on UCI HAR and 82.70% accuracy with 0.128 million parameters on the self-collected dataset. Applying feature-level fusion instead of data-level fusion in the proposed FLF-MSATN improved the model's accuracy by ~3.4% for UCI HAR and ~9.68% for self-collected datasets.

Although the proposed model obtains adequate performance on the self-collected database, it can be further improved with more training samples from different users, as the classifier will have more insight into the unique characteristics of each activity. Thus, we plan to expand our self-collected databases with more volunteers in our future work.

Acknowledgements

This research is supported by Fundamental Research Grant Scheme (FRGS), FRGS/1/2020/ICT02/MMU/02/7. The authors fully acknowledged the Ministry of Higher Education Malaysia (MOHE) for the approved fund which makes this important research viable and effective.

References

Aguileta, A. A., Brena, R. F., Mayora, O., Molino-Minero-re, E., & Trejo, L. A. (2019). Multi-sensor fusion for activity recognition—a survey. *Sensors (Switzerland)*, *19*(17), 1–41. https://doi.org/10.3390/s19173808

Alawneh, L., Mohsen, B., Al-Zinati, M., Shatnawi, A., & Al-Ayyoub, M. (2020, March 1). A Comparison of Unidirectional and Bidirectional LSTM Networks for Human Activity Recognition. *2020 IEEE International Conference on Pervasive Computing and Communications Workshops, PerCom Workshops 2020*. https://doi.org/10.1109/PerComWorkshops48775.2020.9156264

Anguita, D., Ghio, A., Oneto, L., Parra, X., & Reyes-Ortiz, J. L. (2013). A public domain dataset for human activity recognition using smartphones. ESANN 2013 Proceedings, 21st European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. https://api.semanticscholar.org/CorpusID:6975432

Asim, M., Zhu, M., & Javed, M. Y. (2017). CNN based spatio-temporal feature extraction for face anti-spoofing. 2017 2nd International Conference on Image, Vision and Computing, ICIVC 2017, 234–238. https://doi.org/10.1109/ICIVC.2017.7984552

Bai, S., Kolter, J. Z., & Koltun, V. (2018). An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. ArXiv. https://arxiv.org/abs/1803.01271

Ben Mabrouk, A., & Zagrouba, E. (2018). Abnormal behavior recognition for intelligent video surveillance systems: A review. *Expert Systems with Applications*, *91*, 480–491. https://doi.org/10.1016/j.eswa.2017.09.029

Bengio, Y. (2013). Deep learning of representations: Looking forward. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7978 LNAI, 1–37. https://doi.org/10.1007/978-3-642-39593-2_1

Challa, S. K., Kumar, A., & Semwal, V. B. (2022). A multibranch CNN-BiLSTM model for human activity recognition using wearable sensor data. *Visual Computer*, *38*(12), 4095–4109. https://doi.org/10.1007/s00371-021-02283-3

Dua, N., Singh, S. N., & Semwal, V. B. (2021). Multi-input CNN-GRU based human activity recognition using wearable sensors. *Computing*, *103*(7), 1461–1478. https://doi.org/10.1007/s00607-021-00928-8

Garcia, F. A., Ranieri, C. M., & Romero, R. A. F. (2019). Temporal approaches for human activity recognition using inertial sensors. Proceedings - 2019 Latin American Robotics Symposium, 2019 Brazilian Symposium on Robotics and 2019 Workshop on Robotics in Education, LARS/SBR/WRE 2019, 121–125. https://doi.org/10.1109/LARS-SBR-WRE48964.2019.00029

Glorot, X., Bordes, A., & Bengio, Y. (2011). Deep sparse rectifier neural networks. *Journal of Machine Learning Research*, *15*, 315–323. https://api.semanticscholar.org/CorpusID:2239473

Gravina, R., Alinia, P., Ghasemzadeh, H., & Fortino, G. (2017). Multi-sensor fusion in body sensor networks: State-of-the-art and research challenges. *Information Fusion*, *35*, 1339–1351. https://doi.org/10.1016/j.inffus.2016.09.005

Huang, J., Lin, S., Wang, N., Dai, G., Xie, Y., & Zhou, J. (2020). TSE-CNN: A Two-Stage End-to-End CNN for Human Activity Recognition. *IEEE Journal of Biomedical and Health Informatics*, *24*(1), 292–299. https://doi.org/10.1109/JBHI.2019.2909688

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. 32nd International Conference on Machine Learning, ICML 2015, 1, 448–456. http://proceedings.mlr.press/v37/ioffe15.html

Ismail Fawaz, H., Lucas, B., Forestier, G., Pelletier, C., Schmidt, D. F., Weber, J., Webb, G. I., Idoumghar, L., Muller, P. A., & Petitjean, F. (2020). InceptionTime: Finding AlexNet for time series classification. *Data Mining and Knowledge Discovery*, *34*(6), 1936–1962. https://doi.org/10.1007/s10618-020-00710-y

Khan, Z. N., & Ahmad, J. (2021). Attention induced multi-head convolutional neural network for human activity recognition. *Applied Soft Computing*, *110*, 107671. https://doi.org/10.1016/j.asoc.2021.107671

Lea, C., Ren, M. D. F., Reiter, A., & Hager, G. D. (2016). Temporal Convolutional Networks for Action Segmentation and Detection. arXiv. https://arxiv.org/abs/1611.05267

Mercioni, M. A., & Holban, S. (2020). The Most Used Activation Functions: Classic Versus Current. 2020 15th International Conference on Development and Application Systems, DAS 2020 — Proceedings, 141–145. https://doi.org/10.1109/DAS49615.2020.9108942

Mim, T. R., Amatullah, M., Afreen, S., Yousuf, M. A., Uddin, S., Alyami, S. A., Hasan, K. F., & Moni, M. A. (2023). GRU-INC: An inception-attention based approach using GRU for human activity recognition. *Expert Systems with Applications*, *216*, 119419. https://doi.org/10.1016/j.eswa.2022.119419

Minh Dang, L., Min, K., Wang, H., Jalil Piran, M., Hee Lee, C., & Moon, H. (2020). Sensor-based and vision-based human activity recognition: A comprehensive survey. *Pattern Recognition*, *108*. https://doi.org/10.1016/j.patcog.2020.107561

Nair, N., Thomas, C., & Jayagopi, D. B. (2018). Human activity recognition using temporal convolutional network. *ACM International Conference Proceeding Series*. https://doi.org/10.1145/3266157.3266221

Perez-Gamboa, S., Sun, Q., & Zhang, Y. (2021, March 22). Improved Sensor Based Human Activity Recognition via Hybrid Convolutional and Recurrent Neural Networks. INERTIAL 2021 - 8th IEEE International Symposium on Inertial Sensors and Systems, Proceedings. https://doi.org/10.1109/INERTIAL51137.2021.9430460

Raja Sekaran, S., Han, P. Y., & Yin, O. S. (2023). Smartphone-based human activity recognition using lightweight multiheaded temporal convolutional network. *Expert Systems with Applications*, *227*, 120132. https://doi.org/10.1016/j.eswa.2023.120132

Raja Sekaran, S., Pang, Y. H., Ling, G. F., & Yin, O. S. (2022). MSTCN: A multiscale temporal convolutional network for user independent human activity recognition. *F1000Research*, *10*, 1261. https://doi.org/10.12688/f1000research.73175.2

Rasamoelina, A. D., Adjailia, F., & Sincak, P. (2020). A Review of Activation Function for Artificial Neural Network. SAMI 2020 - IEEE 18th World Symposium on Applied Machine Intelligence and Informatics, Proceedings, 281–286. https://doi.org/10.1109/SAMI48414.2020.9108717

Ronald, M., Poulose, A., & Han, D. S. (2021). iSPLInception: An Inception-ResNet Deep Learning Architecture for Human Activity Recognition. *IEEE Access*, *9*, 68985–69001. https://doi.org/10.1109/ACCESS.2021.3078184

Shome, D. (2021). RestHAR: Residual Feature Learning Transformer for Human Activity Recognition from Multi-sensor Data. 2021 8th International Conference on Soft Computing and Machine Intelligence, ISCMI 2021, 181–185. https://doi.org/10.1109/ISCMI53840.2021.9654816

Stein, M., Janetzko, H., Lamprecht, A., Seebacher, D., Schreck, T., Keim, D., & Grossniklaus, M. (2016). From game events to team tactics: Visual analysis of dangerous situations in multi-match data. TISHW 2016 - 1st International Conference on Technology and Innovation in Sports, Health and Wellbeing, Proceedings, Tishw, 1–9. https://doi.org/10.1109/TISHW.2016.7847777

Tan, V. W. S., Ooi, W. X., Chan, Y. F., Tee, C., & Goh, M. K. O. (2024). Vision-Based Gait

Analysis for Neurodegenerative Disorders Detection. *Journal of Informatics and Web Engineering*, *3*(1), 136–154. https://doi.org/10.33093/jiwe.2024.3.1.9

Tang, Y., Teng, Q., Zhang, L., Min, F., & He, J. (2021). Layer-Wise Training Convolutional Neural Networks with Smaller Filters for Human Activity Recognition Using Wearable Sensors. *IEEE Sensors Journal*, *21*(1), 581–592. https://doi.org/10.1109/JSEN.2020.3015521

Teng, Q., Wang, K., Zhang, L., & He, J. (2020). The Layer-Wise Training Convolutional Neural Networks Using Local Loss for Sensor-Based Human Activity Recognition. *IEEE Sensors Journal*, *20*(13), 7265–7274. https://doi.org/10.1109/JSEN.2020.2978772

Ti, Y. F., Connie, T., & Goh, M. K. O. (2023). GenReGait: Gender Recognition using Gait Features. *Journal of Informatics and Web Engineering*, *2*(2), 129–140. https://doi.org/10.33093/jiwe.2023.2.2.10

Ullah, M., Ullah, H., Khan, S. D., & Cheikh, F. A. (2019). Stacked Lstm Network for Human Activity Recognition Using Smartphone Data. 2019 8th European Workshop on Visual Information Processing (EUVIP), Roma, Italy, 2019, 175–180. https://doi.org/10.1109/EUVIP47703.2019.8946180

Wan, J., Li, M., O'Grady, M. J., Gu, X., Alawlaqi, M. A. A. H., & O'Hare, G. M. P. (2021). Time-Bounded Activity Recognition for Ambient Assisted Living. *IEEE Transactions on Emerging Topics in Computing*, *9*(1), 471–483. https://doi.org/10.1109/TETC.2018.2870047

Xia, K., Huang, J., & Wang, H. (2020). LSTM-CNN Architecture for Human Activity Recognition. *IEEE Access*, *8*, 56855–56866. https://doi.org/10.1109/ACCESS.2020.2982225

Xu, C., Chai, D., He, J., Zhang, X., & Duan, S. (2019). InnoHAR: A deep neural network for complex human activity recognition. *IEEE Access*, *7*, 9893–9902. https://doi.org/10.1109/ACCESS.2018.2890675

Yang, J. B., Nguyen, M. N., San, P. P., Li, X. L., & Krishnaswamy, S. (2015). Deep convolutional neural networks on multichannel time series for human activity recognition. IJCAI International Joint Conference on Artificial Intelligence, 2015, 3995–4001. https://api.semanticscholar.org/CorpusID:1605434

Zafar, A., Aamir, M., Mohd Nawi, N., Arshad, A., Riaz, S., Alruban, A., Dutta, A. K., & Almotairi, S. (2022). A Comparison of Pooling Methods for Convolutional Neural Networks. *Applied Sciences (Switzerland)*, *12*(17). https://doi.org/10.3390/app12178643

Zhang, H., Xiao, Z., Wang, J., Li, F., & Szczerbicki, E. (2020). A Novel IoT-Perceptive Human Activity Recognition (HAR) Approach Using Multihead Convolutional Attention. *IEEE Internet of Things Journal*, *7*(2), 1072–1080. https://doi.org/10.1109/JIOT.2019.2949715

Enhancing Book Recommendation Accuracy through User Rating Analysis and Collaborative Filtering Techniques

An Empirical Analysis

Jirat Chongwarin
Khon Kaen University, Thailand

Paiboon Manorom
Khon Kaen University, Thailand

Vispat Chaichuay
Khon Kaen University, Thailand

Tossapon Boongoen
Aberystwyth University, United Kingdom

Chunqiu Li
Beijing Normal University, China

Wirapong Chansanam
Khon Kaen University, Thailand

Abstract: Since online electronic books have become popular, book recommendation systems have been invented and challenged to handle the high demand from users in the digital era. This study aimed to develop and evaluate a book recommendation model using data mining techniques through RapidMiner Studio. The datasets used were comprised of 981,756 user ratings. Before conducting the data analytics, the data was pre-processed to eliminate duplicates and retain only the highest ratings. Collaborative Filtering (CF) techniques, particularly k-Nearest Neighbours (k-NN) and Matrix Factorization (KF), were employed to elicit insightful information for development and to highlight their capabilities in handling enormous datasets. Furthermore, statistical analysis, visualization, elementary modelling, and model combinations were investigated to compare their performance.

To reinforce creditability, modelling techniques and parameter adjustments were integrated to optimize the performance of the algorithms, since the results indicated that different model settings and data partitions impacted the effectiveness of the recommendation system.

https://doi.org/10.18080/jtde.v12n3.976

Additionally, these results demonstrated the potential of hybrid models in improving the accuracy and efficiency of recommendation systems and highlighted the trade-off between algorithmic approaches and dataset characteristics that interplay in optimizing the performance of recommendation systems.

Keywords: collaborative filtering, k-Nearest Neighbours (k-NN), matrix factorization, RapidMiner Studio, data mining techniques.

Introduction

Book Recommendation Systems

Recent recommendation systems (RSs) have significantly improved to provide tailored suggestions to enhance user experience and satisfaction. As a result, this study comprehensively explores various RS methodologies through hybrid systems, which integrated multiple filtering techniques to address the recommendation system challenges, such as scalability, accuracy, and new user integration. The recommendation systems for books have been developed for a decade to improve user (reader) support and precise decision-making for the right target (buyer). Additionally, RSs have been implemented in both commercial business and educational organizations. For instance, in the book recommendation domain, Mathew *et al.* (2016) integrated content-based filtering, collaborative filtering, and association rule mining to resolve the inefficiencies in existing systems by providing more precise book recommendations to align with individual preferences.

The collaborative filtering (CF) technique was introduced to make precise recommendations to boost RSs for higher performance. CF is considered a critical feature in recommendation systems because it can suggest items based on user interaction patterns and operate through user- and item-based approaches by employing similarity metrics to generate recommendations. For example, Rana & Deeba (2019) introduced a method that enhances the accuracy of CF by incorporating Jaccard Similarity to evaluate user overlap between products, improving the effectiveness of recommendations on electronic book (e-book) platforms like Amazon and Goodreads. The latest advances in recommendation systems from this study have significantly improved the RS's ability to provide tailored suggestions that enhance user experience. Another extensive study in recommendation systems was found in You *et al.* (2023), which introduced the Deep Ranking Weighted Multi-Hashing Recommender (DRWMR) System. The system enhances recommendation accuracy and offers explanations for its suggestions, addressing key challenges in CF systems. To confirm this feature, the study

https://doi.org/10.18080/jtde.v12n3.976

by Kumar *et al.* (2023) highlighted the importance of hybrid algorithms that combine CF with content-based filtering to improve recommendation accuracy and personalization to leverage the strengths of these approaches.

Furthermore, Zhao *et al.* (2022) addressed the challenge of sparse data in CF. They proposed a new item-based CF algorithm to handle the vague datasets with Kullback-Leibler (KL) divergence for measuring item similarity and a novel neighbour selection technique. The study improved prediction accuracy to enhance adaptability to sparse data by integrating more comprehensive rating information, as demonstrated by superior performance on benchmark datasets by the Bayesian Personalized Ranking Smart Linear Model (BPRSLIM). The BPRSLIM is a model-based CF algorithm that enhances accuracy by reconstructing the user-item matrix using item-feature information. This model was adopted by the study of Isinkaye (2023), which conducted extensive experiments to reveal significant improvements in recommendation accuracy, especially in precision and normalized Discounted Cumulative Gain, by 30.6% and 22.1%, respectively.

Nonetheless, although RSs have been developed to be compatible with sparse data, some limitations of explicit ratings in sparse data environments remain, such as in the study of Al-Ghuribi *et al.* (2023), which utilized sentiment analysis of user reviews to derive implicit ratings. This study proposed methods considering sentiment degrees and aspect-sentiment word pairs by combining these with explicit ratings to improve CF performance from Amazon and Yelp datasets. This study demonstrated that these approaches significantly enhanced the accuracy of CF rating prediction algorithms.

Collaborative Filtering integrations

Patel & Patidar (2018) conducted a comparative assessment by exploring a hybrid recommendation model for an online book portal that adapts to user demand and book quality. Furthermore, the study of Mankar *et al.* (2023) emphasized the inclusion of the k-Nearest Neighbours (k-NN) machine-learning model within a CF framework to categorize books based on user preferences to boost customer attraction and revenue for online bookstores by enabling users to select the most fitting books. Devika *et al.* (2021) described a CF system to encourage reading habits by matching book selections with user interests through extensive datasets and machine learning techniques. Correspondingly, Bharathi (2019) developed a hybrid RS for video streaming that integrates user feedback with content and CF algorithms to overcome existing limitations and enhance personalization. Apart from integrating machine learning and algorithms, CF was also integrated with social media platforms for higher prediction efficacy, as in the study of Singh & Banerjee (2023), which enhanced e-commerce book recommendations by integrating item-to-item in the CF approaches. Social media interactions and sentiment analysis from Goodreads, Amazon reviews, and Twitter were the main data platforms for data collection and analysis. Based on the study results, the model has achieved high accuracy in predicting user preferences.

Furthermore, when addressing the complex task of predicting multi-label interests in hybrid news recommendation systems, Saravanapriya *et al.* (2022) proposed a deep learning approach using a multi-label Convolutional Neural Network (CNN) to predict user preferences. This feature forecasted the user interests across 15 labels and identified popular news articles by clustering labels mined from social media platforms, such as Facebook and Twitter. The system was integrated with the latest news feeds to generate recommendations to show prediction performance improvement by 5.87%, 12.09%, and 18.49% through machine learning, measured by Support Vector Machine, Decision Tree, and Naïve Bayes, respectively. Sarimehmetoğlu & Erdem (2023) developed a deep learning-based system through YOLOv5 to extract book titles from promotional videos on social media, facilitating efficient video navigation and enhancing text extraction and video analysis. Additionally, the study of Ifada *et al.* (2019) investigated the probabilistic-keyword CF approaches to improve library book recommendations by addressing the sparsity issue in the traditional CF approaches. This study employed circulation and keyword matrices, integrated them into a keyword model, and then adapted a probabilistic technique to generate top-N book recommendations. The results indicated that the performance of the user- and item-based CF methods obtained higher performance than that of other approaches.

Similarly, the study of Tian *et al.* (2019) developed a personalized book recommendation system for college libraries through hybrid algorithms, which integrated CF and content-based methods to improve user-item scoring matrices and clustering to address data sparsity and accuracy. Additionally, Al-Hagery (2020) proposed a hybrid RS that integrates clustering algorithms, user-based CF, and cosine similarity measures to illustrate significant improvements in product recommendation accuracy and reliability. The study of Rao *et al.* (2021) described a hybrid RS that offers personalized book and audiobook suggestions by combining user profiles with machine learning, which corresponded with the study by Raghavendra & Srikantaiah (2022), which addressed the cold-start problem through demographic data to enhance the quality and accuracy of book recommendations.

The recommendation system challenges

Although RSs have been adopted with extensive integration and in various contexts, some challenges in handling user preferences correspond with current customer demand in the digital era. For example, the study of Pasricha & Solanki (2019) identified that influential social network users enhanced book RSs by providing a novel approach to generate accurate recommendations even without extensive purchase history. Similarly, Bharathipriya *et al.* (2020) conducted a collaborative filtering RS with classification and clustering algorithms to categorize similar customers and offer personalized product suggestions to enhance Business-to-Consumer relationships. To overcome these challenges, Hikmatyar (2020) developed a desktop-based book search recommendation system for the University of Struggle library using Python and MySQL, which employed a user-based CF method, which systematically recommended books by analyzing user-profiles and lending patterns, enhancing search efficiency and accuracy. Similarly, Muneer *et al.* (2022) also developed a content-based filtering RS tailored for tour spots that aligns destination suggestions with user preferences, enhancing tourist satisfaction. Furthermore, Munaji & Emanuel (2021) leveraged user ratings and Pearson correlation to gauge user similarity, tailoring restaurant recommendations to user preferences.

To enhance the efficacy ratios, some recent studies employed the integration of machine learning algorithms, such as the study of Chaturvedi *et al.* (2023), which developed a hybrid book recommender system that employs K-means and Gaussian mixture models for clustering, utilizing Term Frequency Inverse Document Frequency (TF-IDF) vectorization for content analysis and achieved high accuracy in recommending books during the COVID-19 pandemic and in the context of e-book availability challenges. Furthermore, Sarma *et al.* (2021) developed a recommendation system that utilizes clustering with K-means, Cosine Distance and Cosine Similarity to suggest books based on similarity, improving search efficacy in traditional user-based systems.

Emerging technologies and frameworks

Ang & Haw (2021) investigated recommender systems for enhancing public understanding and evaluating the accuracy of various methods in predicting item ratings. The study employed Python to create prototypes with graphical interfaces, enabling visualization and assessment of these systems' effectiveness for commercial implementation. In parallel, Li *et al.* (2023) introduced BookGPT, a framework that integrates large language model technology, especially generative pre-trained transformers (GPTs), into book recommendation systems to explore the application of ChatGPT for book rating, user rating, and summary of book recommendations, and to examine the potential of GPTs to enhance traditional models, especially in less data-rich contexts. Additionally, the studies of Khan *et al.* (2017) and Rajalakshmi *et al.* (2024) explored the effectiveness of CF, content-based methods, and hybrid algorithms in recommending products tailored to user interests, especially in the competitive environment of online platforms. The results indicated the dynamic nature and critical role of

https://doi.org/10.18080/jtde.v12n3.976

RSs in refining e-commerce and other online service platforms to confirm that the users received the most relevant and high-quality recommendations.

From the existing studies, we observed the dynamic evolution of recommendation systems emphasize the critical role of CF and hybrid approaches in delivering accurate and personalized recommendations amidst diverse data challenges. As a result, this study aimed to identify a significant gap in the scalability and integration of new users within existing recommendation systems. Despite advances in collaborative filtering (Jain & Vishwakarma, 2017), content-based filtering and hybrid models have struggled to manage the vast amounts of data generated by increasing users and products. This issue was further compounded by the cold-start problem involving new users or items with limited historical data. As a result, the data was complex and could not be integrated effectively into the recommendation algorithms. These difficulties have obstructed the system's ability to provide accurate recommendations from the onset of user interaction, impacting user satisfaction and system efficiency. Thus, these problems attract the researcher's curiosity to develop a more robust RS by integrating advanced machine-learning algorithms and leveraging demographic data to enhance system scalability and improve the accuracy for initial new users and of product recommendations.

Material and Methodology

To develop and evaluate a book recommendation model utilizing data-mining techniques, the researchers employed a comprehensive approach:

a) Dataset Selection: in this study, the researchers employed a well-established dataset pertinent to book recommendations, ensuring it contains a diverse array of books, user ratings, and associated metadata to facilitate a robust analysis.

b) Data Pre-processing: this step involves cleaning the data by removing duplicates, handling missing values, and filtering irrelevant information. The data was transformed and normalized to prepare it for effective analysis.

c) Technique Selection: the researchers selected appropriate data-mining techniques; collaborative filtering was employed in this study.

d) Model Development: this study employed the chosen techniques; the researchers constructed several models to predict user preferences. This involved setting up algorithms in a data-mining tool, RapidMiner, and configuring them with the necessary parameters.

e) Model Training and Validation: the models were trained using a portion of the dataset designated as the training set. The researchers then validated these models on a separate validation set to gauge their accuracy and effectiveness.

https://doi.org/10.18080/jtde.v12n3.976

f) Performance Evaluation: the researchers evaluated each model's performance through accuracy, precision, recall, F1 score, and the Area under the Receiver Operating Characteristics (ROC) curve (AUC). These metrics facilitated determining the model's effectiveness in recommending books that align with user preferences.

g) Model Optimization: based on the performance evaluations, models are fine-tuned and optimized by adjusting parameters, retraining, or incorporating additional features.

Data

The dataset in this study was in rating.csv format and was comprised of rating information for ten thousand widely read books. These ratings were acquired through online platforms, and the scoring system was ranked from 1 to 5, as Foxtrot (2024) documented. The dataset comprises 981,756 rows and 3 attributes; the specifics of each attribute are shown in Table 1.

Table 1. The dataset details

Attribute	Description	Type	Value
Book_id	Book code	Integer	1-10000
User_id	User ID	Integer	1-53424
rating	Score rating by user	Integer	1-5

Upon importing the dataset into RapidMiner Studio, the initial step involves Exploratory Data Analysis to survey and examine the data. Data visualization techniques were employed to facilitate comprehension of the dataset. It was observed that most users rated books between 3 and 5 points, with no ratings lower than 1 point. Subsequently, an analysis was conducted to determine the frequency of ratings per user-book pair. The study revealed that users rated books at least twice and as many as 200 times. Then, plotting this data on a line graph demonstrated high density within the range of 2 to 100 ratings by indicating a common rating frequency among users.

Furthermore, an examination of each book's ratings revealed that the least-rated books garnered 8 ratings, while the most-rated books received 100 ratings. Notably, the line graph depicting this data exhibited dense clusters within the 60 to 100 rating range, which suggested a consistent rating pattern across most books. Overall, these findings provide valuable insights into user behaviour and rating distribution within the dataset for developing the book recommendation model.

Prepare data to be analysed

Data cleaning was prioritised in the preliminary stages of data preparation for analysis to ensure optimal quality. Upon scrutinizing the dataset, it was observed that certain users had submitted multiple ratings for the same book, which contradicts the intended protocol of a single rating per book per user; as a result, 4,487 duplicates were removed and only the highest-rated entry for each book by a user remained, with a total of 979,478 rows. Subsequently, a filtration process was employed for data rating distribution for each user to enhance the accuracy of performance. The analysis revealed that a substantial volume of rating data for each user is essential for the approach to selection. In this study, collaborative filtering techniques considered a user's historical preferences, and data filtering was focused on users with ratings for more than 20 books. To confirm the completion of the data filtration process, the datasets were marked for creating a book recommendation model comprising 720,212 rows, which contained user_id information for 14,612 individuals and book_id data for 9,999 books.

Build and measure the performance of a recommendation model

To develop a recommendation model, the study leveraged a dataset comprising book ratings from users, a common approach for such systems. CF techniques, particularly k-Nearest Neighbour (k-NN) and Matrix Factorization, were implemented, along with a Model Combiner (k-NN + MF) to evaluate their performance through RapidMiner Studio 10.1. The rating.csv dataset was imported via the ReadCSV operator, and the attributes were then defined through the Set Role operator: ratings serve as labels, user_id as user identification, and book_id as item identification. The datasets were split into training data (90%) and test data (10%) using the Split Data operator. The modelling process was launched through the Recommenders extension operator, which was focused on Collaborative Filtering Item Recommendations.

The chosen algorithm was trained on the training dataset to generate a recommendation model. This model was then applied to the test data by providing each user with the top 15 book recommendations. Next, performance evaluation was conducted using the Performance Operator (Item Recommendation) by measuring metrics such as Area Under the ROC Curve (AUC); Precision at 5 (prec@5), which measures the proportion of relevant items among the top 5 recommendations; Precision at 10 (prec@10), which measures the proportion of relevant items among the top 10 recommendations; and Precision at 15 (prec@15), which measures the proportion of relevant items among the top 15 recommendations. Then, the Normalized Discounted Cumulative Gain (NDCG) was used to evaluate the ranking quality of the recommendations through the position of relevant items, with the Mean Average Precision

https://doi.org/10.18080/jtde.v12n3.976

(MAP) used to calculate the average precision for each query, and then the mean across all queries. The parameter modelling process in this study was implemented through the RapidMiner software.

Based on this study, the researchers proposed the development of a recommender model utilizing the k-Nearest Neighbour (k-NN) technique, which was divided into two formats as follows:

a) User-based focuses on users within the system who express their preferences for books in comparable manner, as demonstrated by Bošnjak *et al.* (2011) and Tang & Wen (2015). The user-based k-NN model leverages the concept that users with similar preferences or behaviours can provide valuable recommendations to one another by identifying and analysing similarities. The model can suggest to a user items that similar users had liked or interacted with, to enhance the overall user experience. The development of this model involves several key steps and operators for simulation purposes. Initially, user data was pre-processed to ensure its quality and relevance to calculate the similarity between users using various similarity measures, such as cosine similarity, Pearson correlation, or Euclidean distance. Once the similarities were determined, the k-NN algorithm was applied to identify the nearest neighbours for each user. The model then aggregated the preferences of these neighbours to generate personalized recommendations. Various operators in RapidMiner Studio facilitated the tasks, such as data normalization, similarity computation, model training, and evaluation, throughout this process.

b) The item-based approach took into account books that were consistently similarly preferred by multiple users, as demonstrated by various operators utilized in constructing the model (Bošnjak *et al.*, 2011; Tang & Wen, 2015). The model was constructed using similar operators to those used in the user-based k-NN model, emphasized the process of calculating the similarity between items rather than users, by focusing on item-to-item relationships. The model can recommend items to users based on the similarity of items that they have previously interacted with. The implementation in this study revealed that the practical application of item-based collaborative filtering techniques has highlighted the key steps, such as data pre-processing, similarity calculation, and recommendation generation. This approach was particularly useful in scenarios where item metadata was rich and user interactions were sparse in order to provide a robust alternative to user-based methods.

The study also proposed developing a recommender model utilizing the Matrix Factorization technique (Mihelčić *et al.*, 2012) for incorporating different operators.

https://doi.org/10.18080/jtde.v12n3.976

The matrix factorization technique is popular in recommender systems for predicting user preferences. The technique involves decomposing a large matrix into smaller, more manageable matrices, typically representing user-item interactions. These smaller matrices capture latent factors influencing user preferences, enabling more accurate predictions. It highlighted the process of selecting and tuning these parameters, such as the number of latent factors, regularization terms, and learning rates, which were crucial for improving the model's accuracy and efficiency. This visualization helps us understand how the matrix factorization model operates and how it can be fine-tuned to predict better user preferences based on the hidden patterns in the data.

Results

As in the study objectives, this study aimed to investigate the development of recommendation models utilizing k-NN, Matrix Factorization (MF), and a combined approach of k-NN and MF techniques in order to assess the algorithms' effectiveness. The study results from the implementation of these algorithms are illustrated in the following subsections.

Performance of the k-Nearest Neighbour (k-NN) algorithm

The study explored recommendation modelling employing the k-NN method, which groups data by considering the closest k neighbours. The k-NN operator used the Cosine Similarity method to calculate similarity values. In this method, an angle of 0° indicates high similarity between variables, while an angle of 90° signifies independence or no similarity at all (Bošnjak *et al.*, 2011; Tang & Wen, 2015). The rating.csv dataset imported into the program comprised 720,212 rows, encompassing information on 14,612 users, 9,999 book data entries, and book rating data ranked from 1 to 5 points. The researchers experimented with different values of k to identify the optimal model performance through random selection. The k values of 5, 15, 25, 35, 45, 55, 65, 75, 85, 95, 105, 125, 155, 175, and 205 were evaluated. Subsequently, the dataset was divided into training and testing data in three proportions: 90:10, 80:20, and 75:25. This approach ensured a comprehensive examination of recommendation modelling effectiveness while maintaining computational efficiency.

The user-based approach relies on user data within the system, which assessed preferences for analogous books. The outcomes derived from model processing shown in Table 2 were achieved by partitioning the dataset with a ratio of 90:10.

Table 2 shows the results from random testing to find the best k value for the User k-NN technique through a 90:10 data split. The study revealed a correlation between the k values and the model's performance, with performance tending to improve with higher k values. However, an interesting observation emerged regarding the accuracy at position 5 (prec@5), as there was a positive correlation between accuracy and k values, until reaching Scenario 6 where k was set to 55. In this context, *'position'* refers to the rank of items in the recommendation list, such as the top 5 recommendations. A '*scenario*' refers to a specific experimental setup or configuration used in the study to test the performance of the recommendation model under different conditions. At this point, the results show some repetitions, particularly noticeable in accuracy values at positions 10 and 15, which started repeating from Scenario 6 and Scenario 8, respectively. Similarly, the Normalized Discounted Cumulative Gain (NDCG) value gradually increases until Scenario 14, where it reaches the high point. Conversely, the mean average precision (MAP) shows a consistent increase until Scenario 14, which was slightly decreased in Scenario 15 compared to Scenario 14. Consequently, the researchers ceased randomizing k values after Scenario 15. Thus, the performance results for each k value indicate that Scenario 13, which corresponds to a k value of 155, produces the optimal performance.

Table 2. Optimal k-value results from random 90:10 testing

Metric	Value
AUC	0.812
prec@5	0.078
prec@10	0.068
prec@15	0.066
NDCG	0.327
MAP	0.064

Subsequently, the data splitting ratio was adjusted to 80:20 for the training and testing datasets to ensure consistency with the random k value used in the initial 90:10 split, as shown in Table 3.

Table 3. Optimal k-value results from random 80:20 testing

Metric	Value
AUC	0.889
prec@5	0.198
prec@10	0.146
prec@15	0.121
NDCG	0.437
MAP	0.119

Table 3 shows the results from random testing to find the best k value of the User k-NN technique in the 80:20 data split. The results show a consistent trend across experiments where the 80:20 ratio corresponds to higher k values. Consequently, the model's performance metric is significantly increasing, particularly the Area Under the Curve (AUC), with increasing values of k. Moreover, the accuracy at position 5 (prec@5) shows a positive

https://doi.org/10.18080/jtde.v12n3.976

correlation with the increment of k. However, when considering the results, after reaching Scenario 11 with a k value of 105, subsequent scenarios, such as Scenario 15, show declining results, with accuracy values at positions 10 and 15 showing signs of stagnation. Conversely, the Normalized Discounted Cumulative Gain (NDCG) values show a gradual improvement, culminating in repeated outcomes from Scenario 15 onwards. The Mean Average Precision (MAP) values show a similar pattern of gradual improvement until Scenario 15, after which there are recurring outcomes from previous scenarios. Consequently, the results show the best performance in Scenario 14, where the dataset was divided into an 80:20 ratio and k was set to 175 — this emerged as the preferred configuration.

Table 4 shows the result of the data analysis process where the ratio of partitioning the data into training and testing sets was adjusted from a 90:10 split to a 75:25 split. The adjustment was made using the same random seed to ensure consistency in data sampling between experiments. This practice helps maintain the randomness of the data split while allowing for reproducible results when the experiment is rerun or reviewed. Table 4 shows that by comparing the model's performance under different data splits by using the same random seed across different experiments, variations in model performance can be attributed more directly to changes in the training-testing split, leading to more reliable conclusions about the model's robustness and scalability.

Table 4. Optimal k-value results from random 75:25 testing

Metric	Value
AUC	0.933
prec@5	0.127
prec@10	0.105
prec@15	0.093
NDCG	0.392
MAP	0.082

The k-Nearest Neighbours (k-NN) algorithm was used to assess books based on user ratings. The dataset was partitioned in a 90:10 ratio, with 90% used for training and 10% for testing. Table 5 shows the results for the optimal k value from random 90:10 testing. This table provides key metrics that reflect the model's performance when applying the optimal k value. The k values were balanced between bias and variance to best generalize unseen data.

Table 5. Optimal k-value results from random 90:10 testing

Metric	Value
AUC	0.915
prec@5	0.154
prec@10	0.117
prec@15	0.098

Metric	Value
NDCG	0.396
MAP	0.093

After optimizing the k value for the User k-NN technique, the focus shifted to evaluating the Item k-NN technique. Table 6 shows the random optimal k value evaluation results in the Item k-NN technique. The analysis across various test scenarios indicates that Scenario 2, which sets k at 6, provides this study's best results. These new scenarios were designed to evaluate the impact of different k values on the performance of the Item k-NN technique. The model's performance in this scenario is primarily assessed through the AUC, a key metric for evaluating classifier performance. The AUC shows a notable upward trend from Scenario 1 through Scenario 5 and highlights an improvement in model efficacy as the experiments progressed. However, while the AUC has improved, the other efficiency metrics declined from their original points in subsequent scenarios. This suggests a trade-off between the AUC and other performance indicators: while the model becomes better at distinguishing between classes (as shown in the rising AUC), it might lose efficiency in other areas due to overfitting or increased computational demand. Based on these observations, it can be inferred that using a data splitting ratio of 75:25, Scenario 2 with k = 6 significantly shows the most balanced and effective outcomes. This setting seems to optimize the trade-offs to achieve a high accuracy without excessively compromising on other performance aspects. This conclusion highlights the importance of tuning the k parameter to align with specific data distributions and objectives, demonstrating that even small adjustments can significantly impact the performance of k-NN models.

Table 6. Performance metrics across different k values

Scenario	K	AUC	PREC @5	PREC @10	PREC @15	NDCG	MAP	Execute Time
Scenario1	5	0.743	0.257	0.203	0.170	0.450	0.166	2.02
Scenario2	6	0.755	0.259	0.205	0.172	0.453	0.168	2.19
Scenario3	7	0.765	0.257	0.204	0.172	0.454	0.167	2.28
Scenario4	8	0.773	0.254	0.202	0.171	0.454	0.166	2.47
Scenario5	9	0.780	0.250	0.200	0.169	0.453	0.165	2.57

Performance of the Matrix Factorization model

Matrix Factorization involves decomposing a matrix into smaller matrices, and as a result, the outcomes of these matrices are equal to those of the original matrix. For instance, in the context of user-book ratings, the rating matrix (R) is decomposed into two matrices: the user matrix (X) and the book matrix (Y). The multiplication of these X and Y matrices yields the original rating matrix (R). This decomposition is particularly useful in areas like recommendation systems, signal processing, and data compression. The concept is to simplify a complex matrix into easier to analyze and manipulate components.

In the context of recommendation systems, such as a user-book rating system, matrix factorization assists in predicting unknown ratings by users on a set of books based on the available ratings. The rating matrix, *R*, represents each user's ratings of different books. This matrix is typically sparse, but its meaning might reflect most entries, because not every user has rated every book.

The decomposition involves breaking down the rating matrix into two lower-dimensional matrices, X and Y, often referred to as the user and book matrices, respectively. The user matrix X refers to the latent factors of the users, and the book matrix Y refers to the latent factors of the books. These latent factors include user preferences and book characteristics that cannot be observed directly.

The number of latent factors is much smaller than the number of users or books, and these factors contain the underlying patterns in the data. For instance, in a simple model, each row of X might correspond to a user affinity towards various genres, while each row of Y might correspond to the book genres.

The outcomes of X and the transpose of Y (denoted as X * Y^T, where Y is structured similarly to X) approximates the original matrix R. This outcome provides a prediction of the ratings from a user to a book based on the similar preferences of other users and the attributes of the books. The aim is to fill in the missing entries in the original ratings matrix with these predictions, providing personalized recommendations to users based on their preferences from the data.

Matrix factorization techniques typically were employed in optimization algorithms to investigate the best approximation of R by minimizing the difference between the actual ratings and the predictions. This process is regulated by parameters that control overfitting to assure that the model was generalized adequately to new data through common approaches, including Singular Value Decomposition, Non-negative Matrix Factorization, and Alternating Least Squares, each suited to different data types and applications.

The RapidMiner software incorporates a crucial parameter called '*Iteration*' to control the number of cycles the algorithm runs to reduce discrepancies between predicted outputs and actual data. This parameter is vital, as it directly influences the model's effectiveness by adjusting the weights within the user and book matrices, which is essential for predicting outcomes based on user preferences and item characteristics. The researchers did not adhere to a fixed iteration number in the study, but reached a consensus to choose a method where iteration values were selected randomly within a predetermined range to find the optimal setting for model accuracy instead. The iteration values tested were 5, 15, 25, 35, 45, 55, 65, and 75. This empirical approach tests each configuration to see which yields the best training speed and prediction accuracy balance. To evaluate the effectiveness of these iteration settings, the data was split into training and testing subsets with a 90:10 ratio. This split allows the model to learn from the majority of the data (training set) and then validate the learned patterns on the smaller portion (testing set), which helps assess the model's performance in simulating real-world operations. The results of these experiments are shown in Table 7. Typically, these results would show trends indicating whether higher iterations lead to better performance or if there are diminishing returns after a certain point. This analysis is fundamental in optimizing machine learning models for practical applications to ensure they are both efficient and effective.

Table 7. The random test results of the number of repetitions of the matrix factorization model that divides the dataset in a ratio of 90:10

Scenario	Iteration	AUC	PREC @5	PREC @10	PREC @15	NDCG	MAP	Execute Time
Scenario1	5	0.838	0.017	0.015	0.013	0.196	0.018	2.07
Scenario2	15	0.841	0.017	0.014	0.013	0.196	0.017	1.27
Scenario3	25	0.842	0.017	0.014	0.013	0.195	0.017	1.36
Scenario4	35	0.842	0.017	0.014	0.013	0.196	0.017	1.31
Scenario5	45	0.842	0.017	0.014	0.013	0.196	0.017	2.04
Scenario6	55	0.842	0.017	0.014	0.013	0.196	0.017	2.04
Scenario7	65	0.842	0.017	0.014	0.013	0.196	0.017	2.12
Scenario8	75	0.842	0.017	0.014	0.013	0.196	0.017	2.15

Table 8. Optimal k-value results from random 90:10 testing

Metric	Value
AUC	0.942
prec@5	0.198
prec@10	0.148
prec@15	0.123
NDCG	0.442
MAP	0.120

Table 8 shows the results of testing various iteration values of the matrix factorization model. The initial dataset was divided into a 90:10 ratio, and then was divided into a 75:25 ratio for training and testing (Table 9). The results reveal enhanced performance when setting higher values for latent factors: running the matrix factorization model for 15 iterations and setting the number of latent factors to 100 resulted in the model achieving a peak efficiency value, with an AUC of 0.916. The latent factors refer to the hidden features that represent underlying patterns in the data, and increasing their number can help the model capture more complex relationships. However, other efficiency metrics have declined compared to those in Scenario

6, where a different number of latent factors was used. With the same latent factor set at 100 but only two iterations, exploration into other scenarios mirroring this configuration resulted in Scenarios 3, 9, and 12 showing high model efficiency. The comparative assessment of performance outcomes identified Scenario 12 as yielding the most favourable results, which was achieved by setting the iteration count to 5 and the number of latent factors to 100.

Table 9. The random test results of the number of repetitions of the matrix factorization model that divides the data set in a ratio of 75:25

Scenario	Iter-ation	num factor	AUC	PREC @5	PREC @10	PREC @15	NDCG	MAP	Execute Time
Scenario1	1	25	0.613	0.014	0.012	0.01	0.217	0.009	1
Scenario2	1	50	0.679	0.047	0.037	0.031	0.252	0.027	1.14
Scenario3	1	100	0.714	0.09	0.069	0.057	0.29	0.05	2.03
Scenario4	2	25	0.832	0.101	0.081	0.07	0.32	0.059	1.03
Scenario5	2	50	0.856	0.156	0.123	0.103	0.369	0.093	1.19
Scenario6	2	100	0.873	0.201	0.155	0.13	0.41	0.123	2.45
Scenario7	3	25	0.867	0.103	0.085	0.073	0.33	0.063	1.04
Scenario8	3	50	0.889	0.155	0.123	0.105	0.378	0.095	1.27
Scenario9	3	100	0.899	0.199	0.157	0.133	0.419	0.126	3.24
Scenario10	5	25	0.882	0.09	0.076	0.067	0.326	0.057	1.07
Scenario11	5	50	0.904	0.146	0.118	0.101	0.376	0.091	1.42
Scenario12	5	100	0.911	0.194	0.154	0.132	0.419	0.124	4.5
Scenario13	15	25	0.886	0.081	0.069	0.062	0.32	0.053	1.23
Scenario14	15	50	0.91	0.136	0.112	0.097	0.372	0.087	3
Scenario15	15	100	0.916	0.187	0.149	0.128	0.415	0.12	11.51

Performance of the Model Combiner (k-NN + MF) technique

The results from the integrated three-algorithms model were introduced as the Combiner Model, connecting the strengths of each component. The results show how the Model Combiner operator integrates these models, functioning in parallel and producing outcomes through a weighted average approach. The model's efficacy is evaluated against the default parameters and those parameters optimized through prior experiments.

Table 10 shows that the learning data set was partitioned alongside the test data set at ratios of 90:10, 80:20, and 75:25 to compare the results. Observations indicate that Scenario 1, adhering to default parameters and a 90:10 ratio, obtained notably lower performance metrics, including Accuracy at Positions 5, 10, and 15, compared to other scenarios. Conversely, Scenarios 2 to 4, wherein model variables from previous experiments are optimized, show improved performance in Scenario 2, which obtained the highest AUC value of 0.937, yet lags behind Scenario 4 in accuracy metrics. Conversely, Scenario 4, aligning with parameters derived in previous sections, emerges as the optimal configuration. Therefore, User k-NN is set to k = 205, Item k-NN to k = 5, and Matrix Factorization to iterations = 5 and number of latent factors = 100, alongside a 75:25 learning-test data split.

Table 10. The experimental results of Model Combiner between User k-NN and MF according to the specified experiment

Scenario	Train /Test Data-set	K-user	K-item	Num Factor	Iteration	AUC	PREC@5	PREC@10	PREC@15	NDCG	MAP	Execute Time
1	90/10	80	80	10	30	0.93	0.124	0.087	0.07	0.362	0.136	39.53
2	90/10	205	5	100	5	0.937	0.16	0.115	0.093	0.405	0.174	45.53
3	80/20	205	5	100	5	0.932	0.256	0.196	0.163	0.481	0.194	46.11
4	75/25	205	5	100	5	0.929	0.293	0.229	0.192	0.506	0.201	45.41

In conclusion, Scenario 4 offers the most favourable outcomes, emphasizing the significance of parameter optimization in enhancing model performance, as shown in Table 11.

Table 11. Comparison of performance results across different scenarios

Scenario	Data Split	Parameters	AUC	Accuracy @ Position 5	Accuracy @ Position 10	Accuracy @ Position 15
1	90:10	Default Parameters	Lower	Lower	Lower	Lower
2	80:20	Optimized Parameters from previous experiments	0.937	Higher	Higher	Higher
3	75:25	Optimized Parameters from previous experiments	Higher	Higher	Higher	Higher
4	75:25	User k-NN: k=205, Item k-NN: k=5, Matrix Factorization: Iteration=5, Latent Num Factor=100	Highest	Highest	Highest	Highest

Comparative analysis of model performance

The experimental findings above establish a book recommendation model derived from the goodbooks-10k dataset. Employing the Model Combiner technique yields optimal performance outcomes by integrating three techniques: Item k-NN, User k-NN, and Matrix Factorization. The most efficient settings were ascertained through rigorous parameter configuration across these methodologies. Consequently, the model integration technique emerged as the top performer among all methodologies. This technique entails a 75:25 ratio allocation for learning and test data, alongside parameter specifications for each of the three constituent techniques: the User k-NN technique is parameterized with k = 205, the Item k-NN technique with a k value of 5, and the Matrix Factorization technique with an iteration value of 5 and a hidden factor number of 100. Under these settings, the model achieves the

https://doi.org/10.18080/jtde.v12n3.976

following performance metrics: AUC = 0.942, prec@5 = 0.198, prec@10 = 0.148, prec@15 = 0.123, NDCG = 0.442, and MAP = 0.120, as shown in Table 12. This comprehensive analysis highlighted the efficacy of the model integration technique through parameter optimization across multiple methodologies for superior performance in book recommendation systems.

Table 12. The summary of model performance measure results

Model	AUC	PREC@5	PREC@10	PREC@15	NDCG	MAP
User k-NN	0.889	0.198	0.146	0.121	0.437	0.119
Item k-NN	0.933	0.127	0.105	0.093	0.392	0.082
Matrix Factorization	0.915	0.154	0.117	0.098	0.396	0.093
Model Combiner	0.942	0.198	0.148	0.123	0.442	0.120

Table 12 shows a detailed comparative analysis to evaluate the effectiveness of various recommendation system models, employing the techniques k-NN for both user-based and item-based approaches, matrix factorization, and a hybrid approach that combines these methods. The analysis was based on how well each model predicted user preferences, leveraging the goodbooks-10k dataset—a popular dataset for benchmarking recommendation systems. The key findings highlight that the hybrid model demonstrates superior performance when compared to the individual models (Ngoendee & Charoenruengkit, 2020). This enhancement in performance is attributed to the synergistic effects that arise when these techniques are integrated, and each approach compensates for the weaknesses of the others; for example, matrix factorization can address the sparsity problem in user-item matrices, which is a limitation in k-NN approaches.

Additionally, the study involved experimentally determining the optimal parameters for each recommendation technique. This involved tuning parameters such as the number of neighbours in k-NN models and the regularization and factorization features in matrix factorization models. Identifying these optimal configurations was crucial for enhancing the overall accuracy and efficiency of the recommendation systems. The research indicates that a well-tuned hybrid recommendation model utilizes the strengths of user-based k-NN, item-based k-NN, and matrix factorization to deliver the most effective recommendations, promising optimal outcomes for users and businesses employing this model. This finding is significant, as it guides the development of more robust, accurate, and efficient recommendation systems in practical applications.

Conclusions

In this research, the researchers developed a book recommendation model focusing on Collaborative Filtering techniques, particularly k-NN and Matrix Factorization. The study aimed to enhance model performance by employing data mining techniques and optimizing parameters. The optimal configurations for user-based k-NN, item-based k-NN, and matrix factorization models each produced strong results. The Model Combiner, which integrated all three methods, achieved the highest AUC, 0.942. These results highlight the potential of hybrid models in improving the accuracy and efficiency of recommendation systems. Despite achieving modest accuracy metrics, the study emphasizes the importance of exploring alternative and hybrid methods to address common challenges like scalability and the cold-start problem. Future studies should focus on integrating advanced machine learning techniques, incorporating diverse data sources, and ensuring ethical considerations, such as user privacy and fairness, to enhance recommendation systems.

Acknowledgment

The authors deeply thank Khon Kaen University for its invaluable support. This work was generously funded by a scholarship under the Research Assistant Program at Khon Kaen University, Thailand (Grant No. RA2566-D203). We also sincerely thank Khon Kaen University for facilitating this research through Announcement No. 2580/2563 on "The Criteria for the Acceptance of an Inbound Visiting Scholar from a Foreign Institution or Organization to Khon Kaen University". This support has been crucial to the success of our research project.

References

Al-Ghuribi, S. M., Noah, S. A., & Mohammed, M. A. (2023). An experimental study on the performance of collaborative filtering based on user reviews for large-scale datasets. *PeerJ Computer Science*, *9*, e1525. https://doi.org/10.7717/peerj-cs.1525

Al-Hagery, M. A. (2020). A novel based-approach composed of clustering algorithm & cosine similarity for products recommendation. *International Journal of Education and Information Technologies*, *14*, 133–141. https://doi.org/10.46300/9109.2020.14.16

Ang, J. Y., & Haw, S. C. (2021). Comparative analysis of techniques used in book-based recommender system. In Proceedings of the 5th International Conference on Digital Technology in Education (pp. 87–92). ACM. https://doi.org/10.1145/3488466.3488475

Bharathi, Y. D. (2019). Recommendation system for video streaming websites based on user feedback. *International Journal of Engineering and Advanced Technology*, *8*(6), 1317–1320. https://doi.org/10.35940/ijeat.F8516.088619

Bharathipriya, C., Swathi, B., & Jency, X. F. (2020). Product recommendation framework based on customer review using collaborative filtering techniques. *Journal of Mechanics of Continua and Mathematical Sciences*, (Special Issue 7), 58–71. https://doi.org/10.26782/jmcms.spl.7/2020.02.00004

Bošnjak, M., Antulov-Fantulin, N., Šmuc, T., & Gamberger, D. (2011, June). Constructing recommender systems workflow templates in RapidMiner. In Proceedings of the 2nd RapidMiner Community Meeting and Conference (pp. 101–112). Shaker Verlag. https://www.shaker.eu/Online-Gesamtkatalog-Download/2024.08.10-08.53.51-66.249.66.88-radE74BB.tmp/3-8440-0093-3_INH.PDF

Chaturvedi, A., Subramanian, S. P., Kumar, A., & Mishra, B. (2023). Synergizing collaborative filtering and popularity scores in a machine learning approach for content suggestions. In 2023 7th International Conference on Electronics, Communication and Aerospace Technology (ICECA) (pp. 442–449). IEEE. https://doi.org/10.1109/ICECA58529.2023.10395086

Devika, P. V., Jyothisree, K., Rahul, P. V., Arjun, S., & Narayanan, J. (2021, July). Book recommendation system. In 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT) (pp. 1–5). IEEE. https://doi.org/10.1109/ICCCNT51525.2021.9579647

Foxtrot. (2024). Goodbooks-10k. Kaggle. https://www.kaggle.com/zygmunt/goodbooks-10k

Hikmatyar, M. (2020). Book recommendation system development using user-based collaborative filtering. *Journal of Physics: Conference Series*, *1477*(3), 032024. https://doi.org/10.1088/1742-6596/1477/3/032024

Ifada, N., Susanto, L. R., Saputro, P. A., Nugraha, A. P., Muflikhah, L., Fauzi, M. A., & Adinugroho, S. (2019). Enhancing the performance of library book recommendation system by employing the probabilistic-keyword model on a collaborative filtering approach. *Procedia Computer Science*, *157*, 345–352. https://doi.org/10.1016/j.procs.2019.08.176

Isinkaye, F. O. (2023). Harnessing Item Features to Enhance Recommendation Quality of Collaborative Filtering. *Journal of Applied Intelligent System*, *8*(2),162–172. http://dx.doi.org/10.33633/jais.v8i2.7915

Jain, A., & Vishwakarma, S. K. (2017). Collaborative filtering for movie recommendation using RapidMiner. *International Journal of Computer Applications*, *169*(5), 29–33. https://www.ijcaonline.org/archives/volume169/number6/jain-2017-ijca-914771.pdf

Khan, B. M., Khan, M. T., Malik, A. B., & Ahmad, K. S. (2017). Collaborative filtering based online recommendation systems: A survey. In 2017 International Conference on Information and Communication Technologies (ICICT) (pp. 125–130). IEEE. https://doi.org/10.1109/ICICT.2017.8320176

Kumar, P., Gupta, M. K., Rao, C. R. S., Bhavsingh, M., & Srilakshmi, M. (2023). A Comparative Analysis of Collaborative Filtering Similarity Measurements for Recommendation Systems. *International Journal on Recent and Innovation Trends in Computing and Communication*, *11*(3s), 184–192. https://doi.org/10.17762/ijritcc.v11i3s.6180

Li, Z., Xiao, Q., Liu, Y., Lin, Y., & Luo, F. (2023). BookGPT: A general framework for book recommendation empowered by large language model. *Electronics, 12*(22), 4654. https://doi.org/10.3390/electronics12224654

Mankar, K., Wankhade, P., Bahurupi, S., Jirapure, S., Deshmukh, A., & Thakare, S. (2023). Web based book recommendation system using collaborative filtering. In 2023

International Conference on Emerging Smart Computing and Informatics (ESCI) (pp. 1–6). IEEE. https://doi.org/10.1109/ESCI56872.2023.10099750

Mathew, P., Kuriakose, B., & Hegde, V. (2016). Book recommendation system through content-based and collaborative filtering method. In 2016 International Conference on Data Mining and Advanced Computing (SAPIENCE) (pp. 47–52). IEEE. https://doi.org/10.1109/SAPIENCE.2016.7684166

Mihelčić, M., Antulov-Fantulin, N., Bošnjak, M., & Šmuc, T. (2012). Extending RapidMiner with recommender systems algorithms. Proceedings of the 3rd RapidMiner Community Meeting and Conference (RCOMM 2012). http://bib.irb.hr/datoteka/596976.rcomm2012_recommenders.pdf

Munaji, A. A., & Emanuel, A. W. (2021). Restaurant recommendation system based on user ratings with collaborative filtering. *IOP Conference Series: Materials Science and Engineering*, *1077*(1). https://doi.org/012008. 10.1088/1757-899X/1077/1/012026

Muneer, M., Khan, M. A., Manzoor, S., Naeem, M., & Saeed, H. (2022). Tour spot recommendation system via content-based filtering. In 2022 16th International Conference on Open Source Systems and Technologies (ICOSST) (pp. 1–6). IEEE. https://doi.org/10.1109/ICOSST57195.2022.10016820

Ngoendee, W., & Charoenruengkit, W. (2020). Book recommendation with data mining using RapidMiner [Master's project, Srinakharinwirot University]. Institutional Repository. http://ir-ithesis.swu.ac.th/dspace/bitstream/123456789/1232/1/gs591130027.pdf

Pasricha, H., & Solanki, S. (2019). A new approach for book recommendation using opinion leader mining. In Proceedings of the International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT 2018) (pp. 501–515). Springer Singapore. https://doi.org/10.1007/978-981-13-5802-9_46

Patel, D., & Patidar, H. (2018). Hybrid recommendation solution for online book portal. *International Journal for Research in Applied Science and Engineering Technology*, *6*(5), 1367–1373. http://doi.org/10.22214/ijraset.2018.5225

Raghavendra, C. K., & Srikantaiah, K. C. (2022). Switching hybrid model for personalized recommendations by combining users demographic information. *Journal of Theoretical and Applied Information Technology*, *100*(3), 825–835. http://www.jatit.org/volumes/Vol100No3/20Vol100No3.pdf

Rajalakshmi, S., Indumathi, G., Elias, A., & Priya, G. S. (2024). Personalized Online Book Recommendation System Using Hybrid Machine Learning Techniques. *International Journal of Intelligent Systems and Applications in Engineering*, *12*(15s), 39–46. https://www.ijisae.org/index.php/IJISAE/article/view/4712

Rana, A., & Deeba, K. (2019). Online book recommendation system using collaborative filtering (with Jaccard similarity). *Journal of Physics: Conference Series*, *1362*(1). https://doi.org/012130. 10.1088/1742-6596/1362/1/012130

Rao, B., Bhargava, P., Panchal, D., Rane, S., & Lalwani, P. (2021). Book recommendation system with relevant text audiobook generation. *International Journal of Creative Research Thoughts (IJCRT)*, *42*(1), 398–404. https://ijcrt.org/papers/IJCRT2107170.pdf

Saravanapriya, M., Senthilkumar, R., & Saktheeswaran, J. (2022, July). Multi-label Convolution Neural Network for Personalized News Recommendation based on Social Media Mining. *Journal of Scientific and Industrial Research (JSIR)*, *81*, 785–797. http://op.niscair.res.in/index.php/JSIR/article/download/46261/465480990

Sarimehmetoğlu, B., & Erdem, H. (2023). Extracting Book Titles From Book Recommendation Videos Using a Deep Learning Approach. *MANAS Journal of Engineering*, *11*(2), 229–234. https://doi.org/10.51354/mjen.1369636

Sarma, D., Mittra, T., & Shahadat, M. (2021). Personalized book recommendation system using machine learning algorithm. *International Journal of Advanced Computer Science and Applications*, *12*(6), 212–219. https://doi.org/10.14569/IJACSA.2021.0120126

Singh, K. K., & Banerjee, I. (2023). Integrated personalized book recommendation using social media analysis. *Parikalpana KIIT Journal of Management*, *19*(1), 106–123. http://doi.org/10.23862/kiit-parikalpana/2023/v19/i1/220834

Tang, Z., & Wen, Z. (2015). Recommendation system based on collaborative filtering in RapidMiner. *Computer modelling & new technologies*, *18*(11), 1004–1008. http://www.cmnt.lv/upload-files/ns_19art162.pdf

Tian, Y., Zheng, B., Wang, Y., Zhang, Y., & Wu, Q. (2019). College library personalized recommendation system based on hybrid recommendation algorithm. *Procedia CIRP*, *83*, 490–494. https://doi.org/10.1016/j.procir.2019.04.126

You, Z., Hu, H., Wang, Y., Xue, J., & Yi, X. (2023). Improved Hybrid Collaborative Filtering Algorithm Based on Spark Platform. *Wuhan University Journal of Natural Sciences*, *28*(5), 451–460. https://doi.org/10.1051/wujns/2023285451

Zhao, W., Tian, H., Wu, Y., Cui, Z., & Feng, T. (2022). A New Item-Based Collaborative Filtering Algorithm to Improve the Accuracy of Prediction in Sparse Data. *International Journal of Computational Intelligence Systems*, *15*, a15 https://doi.org/10.1007/s44196-022-00068-7

LongT5Rank: A Novel Integrated Hybrid Approach for Text Summarisation

Agatha Jin Jin Lau
Department of Mathematical and Data Science, Tunku Abdul Rahman University of Management and Technology, Malaysia

Chi Wee Tan
Department of Computer Science and Embedded Systems, Tunku Abdul Rahman University of Management and Technology

Abstract: Text summarisation reduces text length while retaining important information, helping individuals, especially students, in managing information overload during research or assignments. However, existing text summarisation methods often lose important details, generate irrelevant or redundant sentences, or produce incoherent summaries. This study introduces a hybrid approach, LongT5Rank (coined in this study), which combines TextRank, an extractive summarisation algorithm, with LongT5, an abstractive summarisation algorithm, to automate the summarisation process. TextRank utilizes GloVe, a pre-trained word embedding model, and PageRank, a graph-based ranking algorithm, to select representative sentences. LongT5, an encoder-decoder transformer model, condenses extracted sentences into a concise and coherent summary, handling input sequences up to 16,384 tokens, for long-range sequence-to-sequence tasks. The LongT5Rank approach has shown significant achievements, including a minimum 60% compression rate, a minimum 0.6 semantic textual similarity score, and an improved F-measure compared to employing TextRank alone. Furthermore, it received positive feedback from Human Level Performance (HLP), underlining the importance of evaluating results directly from human users. This emphasizes the belief that the performance of the proposed solution should be assessed natively by humans. By combining both extractive and abstractive methods, LongT5Rank excels in generating accurate and coherent summaries.

Keywords: Hybrid approach, LongT5 model, Semantic Textual Similarity, TextRank, Text Summarisation

Introduction

Text summarisation aims to condense lengthy articles into concise summaries by extracting key information and preserving the overall meaning. This process facilitates efficient

https://doi.org/10.18080/jtde.v12n3.977

comprehension of long articles, enabling readers to grasp the main points without sacrificing important details (Mishra, 2022). The summary can be extractive, in which a subset of sentences from the original text is picked for the summary, or abstractive, in which totally new sentences are produced and inserted in the summary.

Extractive summarisation is the process of extracting the most essential sentences or phrases from the source text and combining them to form a summary (Khor *et al.*, 2022). This strategy is commonly utilized in news articles and other sorts of content where the essential points are already conveyed clearly in the text.

Abstractive summarisation is the process of creating new phrases that encapsulate the substance of the original material. This strategy is more difficult and necessitates a better comprehension of the text and its context (Gambhir & Gupta, 2016).

A hybrid approach in text summarisation combines extractive and abstractive approaches to provide a summary. This approach creates an extractive summary from numerous input documents before generating an abstractive summary (Ghadimi & Beigy, 2022).

TextRank is a text summarisation method that analyses word order and context in a document. It is used to extract the most significant information from a source text and compress it into a shorter version while keeping the original text's meaning intact (Zaware *et al.*, 2021).

LongT5 is a modified version of the T5 model that can handle longer input sequences efficiently. It uses sparse attention methods, such as windowed attention and global-local variation, to improve its capacity to digest lengthy inputs. LongT5 has been proven to perform well across a variety of text summarization datasets, surpassing competing models in text creation tasks (Rusnachenko, 2024).

Despite advances in text summarization approaches, there is still a challenge in producing high-quality summaries that properly balance conciseness with coherence, particularly for long documents. Although TextRank is efficient in identifying important sentences, it often struggles to generate a coherent and fluent summary. In contrast, some purely abstractive summarization approaches are able to produce a more coherent and fluent summary than extractive text summarization approaches by rephrasing and restructuring the original content. However, these methods often struggle with maintaining the original meaning when rephrasing the contents.

The motivation for using LongT5 is its efficient attention mechanisms in handling long input sentences. This makes it suitable for processing longer text documents. LongT5 employs sparse attention techniques, such as Local Attention and Transient Global Attention, to enhance its ability to process long inputs. This allows it to retain context and coherence throughout lengthy texts, addressing one of the significant limitations of purely abstractive methods.

Using the combination of the TextRank algorithm and the LongT5 model, which acts as a hybrid approach-based text summariser, provides a solution to this problem by automatically generating concise summaries of longer texts. The TextRank algorithm is an extractive summarisation, while the LongT5 model is an abstractive summarisation.

By combining these two techniques, the text summariser, which we call "LongT5Rank", can potentially utilise the strengths of both extractive and abstractive methods, leading to a more robust and nuanced summarisation process. The TextRank algorithm helps in identifying the most important sentences of the text, which are then used as input for the LongT5 model, resulting in generating a more concise and coherent summary. This approach can help to ensure that the summary captures all the essential meaning of the original text, while also being easy to read and understand, even if it uses different words or phrases.

Literature Review

Automatic Text Summarisation (ATS) is a method for automatically creating a summary of a given text or document. With the exponential expansion of textual material on the Internet and in other archives, manual text summarising has become infeasible and time-consuming. ATS approaches attempt to solve this issue by picking the most significant sentences or producing new sentences that communicate the same information as the original text in a more succinct manner (El-Kassas *et al.*, 2020).

Extractive text summarisation

Extractive summarisation, proposed by El-Kassas *et al.* (2020), is a text summarisation approach that includes extracting relevant lines or phrases from the source text and combining them to generate a summary. The primary idea behind extractive summarisation is to select the most essential phrases in a text and utilize them to construct a summary that contains the key information.

Extractive summarisation, a widely studied approach in the field of natural language processing, involves a series of essential steps proposed by various researchers. Scholars such as Hernandez-Castaneda *et al.* (2020) have contributed to the advancement of extractive summarisation techniques. First, the text is pre-processed to remove extraneous information. Following that, sentences are rated based on their relevance to the document's main topics, using approaches such as TF-IDF, TextRank, or LSA (see below). The most significant sentences are then chosen and merged to make the summary, with the summary length predetermined or dynamically generated. Finally, an evaluation metric, such as ROUGE,

https://doi.org/10.18080/jtde.v12n3.977

which was proposed by researchers, is used to analyse the quality of the generated summary. This approach serves as a fundamental framework for extractive summarisation, allowing the development of succinct summaries by extracting important sentences from the original content.

Several techniques for extractive summarisation have been developed, each with its own set of strengths and shortcomings. Below are some of the techniques used in extractive text summarisation:

TF-IDF

Zaware *et al.* (2021) proposed an approach that utilizes the TF-IDF (Term Frequency-Inverse Document Frequency) algorithm for automatic text summarisation. The TF-IDF technique is a statistical approach for determining the importance of words in a document. It gives a complete measure of word significance by computing the product of two variables, term frequency (TF) and inverse document frequency (IDF).

Term frequency (TF) is the frequency with which a term appears in a document, representing its relative importance within that textual context. If a term appears frequently, its TF value rises, indicating its importance in the document:

$$TF = \frac{\text{number of times the term appears in the document}}{\text{total number of terms in the document}} \quad (1)$$

In contrast, inverse document frequency (IDF) assesses the significance of a term across an entire collection of documents. Words that appear only seldom across the corpus have higher IDF values because they are thought to be more important in transmitting meaningful information.

$$IDF = log(\frac{\text{number of the documents in the corpus}}{\text{number of documents in the corpus contain the term}}) \quad (2)$$

TF-IDF combines the TF and IDF values together to assess the significance of each word in the document. The algorithm then arranges the words in descending order depending on their ratings. The summary is created by selecting and combining the top-ranked terms. The TF-IDF algorithm successfully captures the most important material from the original text by prioritizing terms with higher scores, allowing for the development of informative and short summaries.

$$TF\text{-}IDF = TF * IDF \quad (3)$$

The TF-IDF algorithm, highlighted in the research paper by Zaware, offers several strengths. Firstly, it is a simple and effective method for identifying key words in a document. It is able

https://doi.org/10.18080/jtde.v12n3.977

to discover the words that are most significant to a document by computing the TF-IDF score for each word in that document. This can be beneficial for a range of natural language processing applications, such as text categorization, clustering, and information retrieval.

In the context of identifying gaps in existing knowledge, it is essential to highlight the limitations of the TF-IDF algorithm when applied to text summarisation of long articles for students. The limitations of the technique have been well-documented in the research paper by Zaware *et al.* (2021). As mentioned earlier, the algorithm does not examine the order of words in a document, nor does it analyse the context of words. This can be a limitation for text summarisation when word order and context are critical for producing a brief and accurate summary. Furthermore, the algorithm may be biased against lengthier articles, making it harder to construct a summary of the most significant information in a long article. As a result, the TF-IDF method is less ideal for text summarisation of large articles for students due to its constraints.

Latent Semantic Analysis (LSA)

According to Jayan & Govindaru (2022), Latent Semantic Analysis (LSA) is one of the text summarisation algorithms for Malayalam documents. They investigated LSA, which is a mathematical approach that analyses correlations between a set of documents and the terms they include, by providing a set of ideas connected to the documents and terms. In their study, LSA is used to extract the most essential phrases from Malayalam texts by analysing the connections between the words in the documents. The approach generates a matrix of term-document frequency and uses Singular Value Decomposition (SVD) to minimise the dimensionality of the matrix. The reduced matrix is then used to compute the similarity between the statements and rank them in order of relevance.

The researchers' work demonstrates several notable strengths, including its ability to capture the underlying meaning of the text and its effectiveness in dealing with synonymy and polysemy. By analysing the relationships between the words in the documents, LSA can capture the underlying meaning of the text. This implies it can recognise the major ideas and concepts in a document, even if they are not clearly mentioned. Besides, LSA is effective at dealing with the frequent difficulties of synonymy and polysemy in natural language processing. Synonymy refers to when various words have the same meaning, and polysemy refers to when the same word has many meanings. LSA can recognise the connections between these words and concepts and extract their underlying meaning. In addition, Jayan & Govindaru (2022) highlighted that LSA can be used for both single-document and multi-document summarisation, which implies that it can summarize a single document or multiple documents on the same topic.

https://doi.org/10.18080/jtde.v12n3.977

Despite the significant contributions of this study, LSA exhibits limitations when applied to summarizing long articles. Firstly, LSA requires a vast quantity of data in order to generate reliable findings. This implies that it may not be successful in summarizing lengthy articles with a lot of information. The more information there is, the more difficult it gets for LSA to select the most significant ideas and concepts. Moreover, LSA may be ineffective when dealing with uncommon or unknown words. This implies that, if a long article contains numerous uncommon technical phrases or jargon, LSA may be unable to capture their underlying meaning and significance. Hence, due to its restrictions, it may not be the greatest choice for summarizing large articles for students.

TextRank

The TextRank algorithm is a graph-based ranking system used in natural language processing (NLP) for text summarisation. It is based on Google's PageRank algorithm, which is used to rank websites. Sentences in a text are represented as nodes in a graph in the TextRank algorithm, and the edges between the nodes represent the similarity of the sentences. The algorithm then determines the significance of each sentence based on its position in the network and the significance of the sentences to which it is linked. The most important sentences are then chosen to create a text summary (Joshi, 2023).

In the study of Luo *et al.* (2022), they proposed the use of the TextRank algorithm for Chinese text summarisation. Different variations of the algorithms have been experimented with by them, such as different similarity measurements and weighting techniques, to determine the ideal configuration for summarisation. Their findings suggest that TextRank is an excellent approach for Chinese text summarisation, and that the use of cosine similarity and inverse document frequency (IDF) weighting generates the best results. They also compared their results to those of LSA, LexRank, and others, and found that TextRank beats them in terms of ROUGE scores, as well as human evaluation. Bichi *et al.* (2023) also proposed a modified version of the TextRank algorithm that employs modified inverse sentence frequency-cosine similarity to assign different weightage to different words in the sentence, whereas traditional cosine similarity treats the words equally.

In a recent study by Khor, Tan & Lim (Khor *et al.*, 2022), the TextRank algorithm was used to summarise Amazon food reviews. The goal was to condense lengthy writings into concise summaries by selecting the top five most representative reviews. The process includes using TextRank and GloVe word embeddings to identify key sentences based on their importance and relevance. Each word is mapped to a vector representation using GloVe pre-trained word embeddings, allowing for easier semantic analysis. Furthermore, the PageRank algorithm is important in identifying the significance of sentences, which helps with the making of useful summaries. However, issues, like high computation times and the possible loss of detail in the

https://doi.org/10.18080/jtde.v12n3.977

summaries, are mentioned. Despite these limitations, extractive summarisation shows potential in condensing enormous amounts of text into relevant summaries, with room for future development and investigation.

Abstractive text summarisation

Abstractive summarisation is an automated process for generating summaries, which involves rewriting and reformulating text from the original document, similar to how a human would summarize. It focuses on discovering essential information, evaluating the context, and developing a new set of phrases for the summary. Dewda (2022) considers this method more difficult than extractive summarising, since it needs the extraction of important information as well as the generation of coherent text.

Several techniques for abstractive summarisation have been developed, each with its own set of strengths and shortcomings. Below are some of the techniques used in abstractive text summarisation.

Recurrent Neural Network (RNN)

Abujar *et al.* (2020) proposed an approach for developing abstractive text summarization in the Bengali language using a recurrent neural network (RNN) with long short-term memory (LSTM). The paper aimed to generate summaries that resemble human-written text. RNNs, particularly effective in capturing temporal dependencies and patterns in sequential data, are well-suited for tasks such as language modelling, machine translation, and speech recognition. LSTMs, which address the vanishing gradient problem, enhance the RNN’s ability to model long-term dependencies and have been successfully applied in various natural language processing tasks, including abstractive text summarization. By using an RNN with LSTM, the approach can capture the contextual information and dependencies in the text, leading to more accurate and coherent summaries.

However, in this study, the paper mentions a few limitations. This Bengali Abstractive Text Summarisation may require significant computational resources and time for training and fine-tuning the RNN-LSTM model, which can be a limitation in practical applications. Furthermore, the effectiveness of the approach may depend on the quality of the initial training data and the ability of the model to capture the nuances and complexities of the Bengali language. Hence, due to the model struggling with understanding specialized or technical language, the performance of the approach may vary depending on the specific domain or genre of the text being summarised.

https://doi.org/10.18080/jtde.v12n3.977

Table 1. Comparison of Text Summarisation Algorithms

Algorithm	Strengths	Limitations
TF-IDF	-Simple and effective way to identify important words in a document. -Applicable to various natural language processing tasks, such as text classification, clustering, and information retrieval. -Computationally efficient and easy to implement.	-Does not take into account the order of words in a document. -May be biased towards lengthier articles.
LSA	-Captures underlying meaning of the text. -Effective in dealing with synonymy and polysemy. -Can be used for single-document and multi-document summarisation.	-Requires a vast amount of data for reliable results. -Less suitable for summarizing lengthy articles with abundant information. -May struggle with uncommon or unknown words.
TextRank	-Can be easily implemented. -Does not require any training data or prior knowledge. -Can handle single and multi-document summarisation.	-May produce redundant or incomplete summaries. -May require some parameter tuning to achieve optimal performance.
RNN	-Captures temporal dependencies effectively. -Suitable for various sequential tasks. - LSTM addresses vanishing gradient problems. -Captures context for accurate summaries.	-Requires significant computational resources. -Effectiveness depends on training data quality. -May struggle with specialised language.
LongT5	-Transformer-based, achieves state-of-the-art results. -Introduces efficient attention mechanisms. -Outperforms original T5 models.	-Limited to cased English SentencePiece vocabulary model. -Challenges with lengthy input sequences. -TGlobal attention may cause a modest performance reduction. -Performance influenced by input text quality and biases.

LongT5

Transformer-based neural models are widely used for text summarization. The LongT5 model, introduced in 2022, is an extension of the T5 model that employs sparse attention techniques, like Local Attention and Transient Global Attention, to handle long input sequences more effectively (Guo *et al.*, 2022). LongT5 enhances the performance of transformer-based models by scaling both input size and model size.

LongT5 integrates attention mechanisms from long-input transformers and pre-training strategies from summarization tasks. It introduces the Transient Global (TGlobal) attention mechanism, which mimics local/global attention without additional inputs. LongT5 has achieved state-of-the-art results in summarization and question-answering tasks, outperforming the original T5 models. It is implemented with JAX and the Flax-former library and comes in three sizes: base, large, and xl.

However, LongT5 has some limitations. It is restricted to the English SentencePiece vocab. model and can struggle with extremely long input sequences due to computational constraints. The TGlobal attention method, while efficient, may slightly underperform compared to full attention models. Additionally, the quality of the input text and biases in the training data can influence the model's output.

Refer to Table 1 for a summary of the comparison of the text summarisation algorithms that have been discussed above.

Proposed Method

This section shows the framework of TextRank with LongT5 algorithm in Figure 1 and a detailed description for each step is discussed.

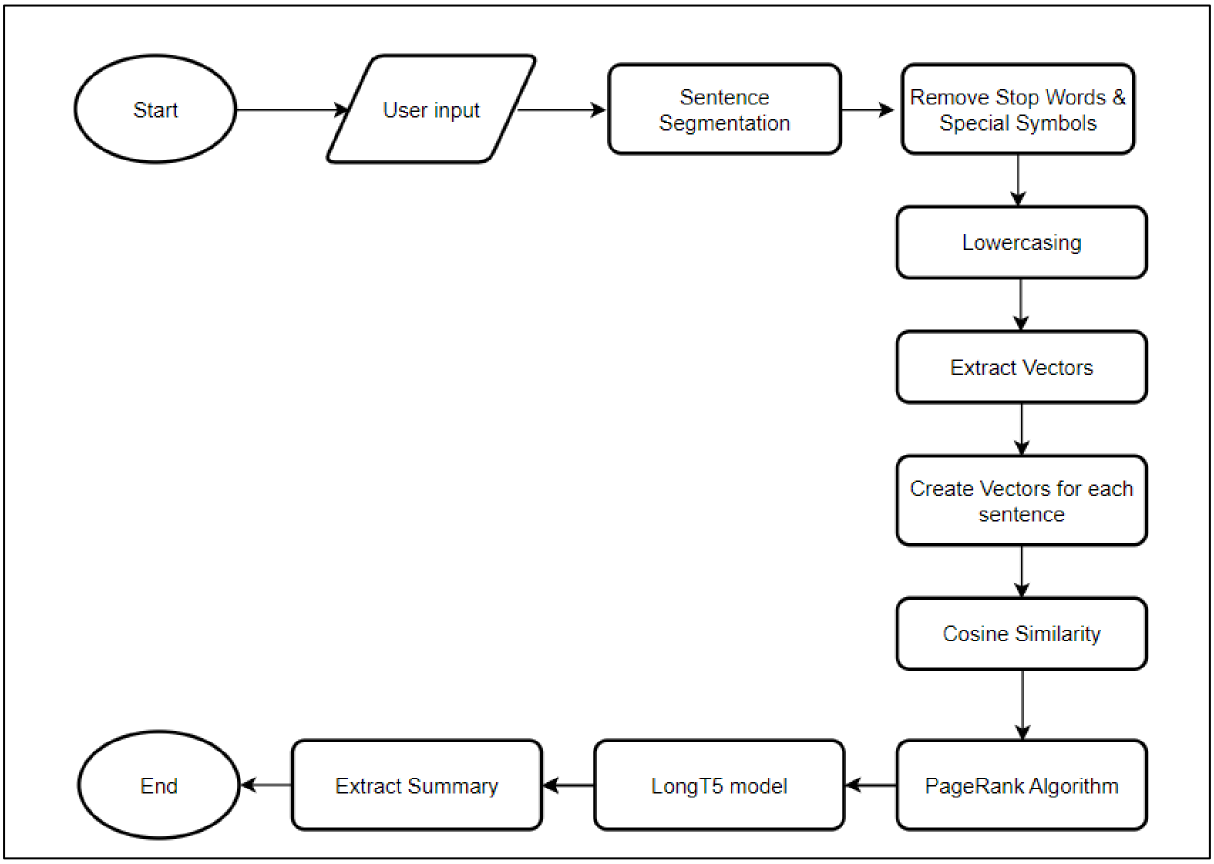


Figure 1. LongT5Rank Model flowchart of Text Summarisation

https://doi.org/10.18080/jtde.v12n3.977

Description of dataset

This project used several news articles which are obtained from the CNN/Daily Mail dataset to perform text summarisation. The dataset comprises articles extracted from CNN and Daily Mail news sources. Furthermore, each of the articles comes with corresponding golden summaries. Because of its excellent quality and broad coverage of subjects, this dataset is commonly used in the field of NLP for training and evaluating summarisation algorithms.

The first phase of the text summarisation process involves directly gathering textual content from user inputs.

Sentence segmentation

After gathering the data, the following step is to segment it into individual sentences as the input text given by the user may consist of a few sentences. When the input text contains question marks, exclamation marks, or full stops, it is segmented.

Data cleaning

Data cleaning is an important step in text summarisation as the text may be noisy and this will greatly impact the performance of text summarising. Hence, text pre-processing should be carried out before summarising the text.

The data-cleaning step involves pre-processing the text to prepare it for summarization. It starts with replacing any non-alphanumeric characters, such as symbols, punctuation and non-space characters, with a space. Then, it converts entire sentences into lowercase. Finally, the pre-processed text can be encoded into a format that the summarisation model can understand, such as word embeddings.

TextRank algorithm

The PageRank algorithm has been adapted into the TextRank algorithm (Joshi, 2023). Unlike the PageRank algorithm, which is used to rank web pages, it is used to rank sentences instead.

Extract word vectors

Before applying the TextRank algorithm, the first step is to extract the word vectors for the words in the text. This technique converts each word into a real-valued vector that is called Word Embedding. In this project, GloVe word embeddings are used to generate word vectors for each word in the clean input text. GloVe is an unsupervised learning technique that trains and creates vector representations for words by aggregating global word-word co-occurrence information from a corpus. A corpus is a large collection of texts that are used for linguistic analysis and natural language processing (Chawla, 2020). The purpose of GloVe is to represent words in a continuous vector space, where similar words are closer to each other in the vector space. Hence, it can help to capture semantic and syntactic relationships between words.

Cosine similarity score

Next, the cosine similarity approach is used to find similarities between the sentences. The cosine similarity measure is mainly used to focus on the common words between sentences and measure the similarity between sentences by calculating the cosine of the angle between the vectors (Krishnan, 2022). Thus, a cosine matrix is generated to store the cosine similarity values; and a value of 0 indicates that the sentences are dissimilar to each other. The cosine matrix is used as the input for the PageRank algorithm, which assigns a score to each sentence based on its importance and relevance.

PageRank algorithm

The similarity matrix is then converted into a graph, where nodes represent sentences and edges represent the relationships between sentences based on their shared topics and semantics. In TextRank, nodes represent sentences in a text document, while edges indicate the relationships between sentences. Then, the PageRank algorithm can be used to compute the importance of sentences in a text by treating the sentences as nodes in a graph and the connections between them as edges. Each sentence will be assigned a score depending on the number and quality of connections it has with other sentences. This score can then be used to determine the importance and relevance of each sentence when forming a summary of the text. In TextRank, the sentence scores will be sorted in descending order, where the sentence with the highest score is the most important sentence in the text. In this project, we will develop a ranked TextRank summary by choosing the top N sentences. The graph in Figure 2 is the graph generated by the "networkx" library.

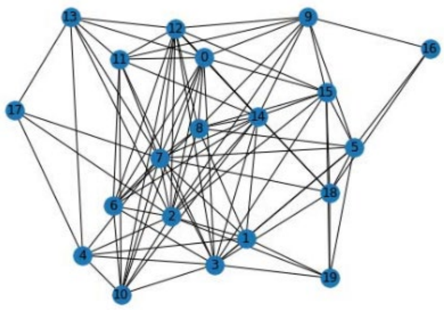


Figure 2. A graph was created based on the similarity scores and was subsequently processed using the PageRank function to generate scores

LongT5

The crucial and final step of this proposed system involves the LongT5 models, which play an important role in the text summarising process. First, the model encodes each phrase from the input text and breaks it down into essential components to understand the meaning of the

https://doi.org/10.18080/jtde.v12n3.977

sentence. Using an attention mechanism, LongT5 focuses on each word in the sentence to learn more about its meaning. The model then combines this understanding of individual words and sentences with the key insights gained from the text and decodes the information to construct a concise and coherent summary.

Extract summary

The TextRank algorithm ranks the sentences based on their importance and relevance. The top N ranked sentences are then fed into the LongT5 model, which refines and structures the content. This fusion process generates a concise and coherent summary that effectively captures the key information from the original text.

Results and Discussion

The evaluation methods that have been used in this project are Semantic Textual Similarity, ROUGE-1 and the Human Level Performance (HLP).

Semantic textual similarity

The first approach was using Semantic Textual Similarity (STS). This measures how well the summary has captured the essence of the original text by assigning a score, where a higher score indicates it has done a good job of preserving the original meaning, even if it uses different words or phrases (Majumder *et al.*, 2016).

Table 2 shows the results for the TextRank summariser using semantic textual similarity as a measure. In this table, the numbers “0.4” and “0.7” represent compression ratios, indicating the proportion of the original text retained in the summary. A compression ratio of 0.4 means that the summary is 40% of the original text's length, while a ratio of 0.7 means the summary retains 70% of the original length.

From Table 2, it can be observed that the TextRank method achieves the highest STS scores in two of the three samples but performs lower on Sample 2. This suggests that, among the methods evaluated, TextRank is most effective at preserving the semantic meaning of the original text in its summaries. However, it is important to note that the TextRank summaries are not cohesive as they extract key sentences from the original source text by ranking. The key sentences that TextRank algorithm selects are likely to be the most semantically rich and representative of the overall text. Therefore, even if the summary ratio changes, as long as these key sentences are included in the summary, the STS score with respect to the original text might remain the same.

https://doi.org/10.18080/jtde.v12n3.977

Table 2. Results of summaries with Semantic Textual Similarity *(higher scores indicate better results)*

Sample	TextRank	LongT5Rank
Sample 1		
0.4	0.896359920501709 ✓	0.6444788575172424
0.7	0.896359920501709 ✓	0.7373626232147217
Sample 2		
0.4	0.5569220185279846	0.603785514831543 ✓
0.7	0.5569220185279846	0.6212045550346375 ✓
Sample 3		
0.4	0.8429946303367615 ✓	0.5808538198471069
0.7	0.8429946303367615 ✓	0.755942165851593

ROUGE

ROUGE (Recall-Oriented Understudy for Gisting Evaluation) is the most often used collection of standards to evaluate machine-generated summaries (Priyanka, 2022). The ROUGE score is commonly employed in text summarization tasks to assess how well a machine-generated summary captures the important content of a longer text. It helps in determining the similarity between the reference and the summary produced. In other words, it helps in measuring how good the summary generated is by comparing it to a summary written by a human (Santhosh, 2023).

The ROUGE metric encompasses precision, recall and F-measure, where a higher precision suggests that the summariser may be more accurate in selecting relevant sentences for the summary, and a higher recall indicating that the summary captures more relevant sentences. A high F-measure score (F1) means that the summariser is both accurate in selecting relevant sentences (precision) and effective in capturing a significant portion of the relevant information present in the original text (recall). The formula for ROUGE-N is as follows:

$$\text{ROUGE-N} = \frac{\sum_{S\in\{ReferemceSummaries\}} \sum_{gram_n \in S} Count_{match}(gram_n)}{\sum_{S\in\{ReferenceSummaries\}} \sum_{gram_n \in S} Count(gram_n)} \quad (4)$$

ROUGE-N evaluates the similarity between a machine-generated summary and human-generated reference summaries by comparing the frequency of n-grams (e.g., unigrams, bigrams) in both texts. The score is calculated by counting matching n-grams in the candidate summary and dividing the sum by the total number of n-grams in the reference summaries. A higher score indicates better alignment with the reference summaries. ROUGE-1, which measures the overlap of single words, has been used in this project. The results for the summariser using ROUGE-1 as evaluation metric have been shown in Table 3.

From Table 3, the F-measure, which is the harmonic mean of precision and recall, varies across samples. A higher F-measure indicates a better balance between precision and recall, suggesting a more accurate summary compared to the human-written summary. Based on the results, the text summarizer with LongT5Rank generates a higher F-measure score compared to the text summarizer with TextRank alone, except in the case of Sample 1, where TextRank performs better.

Table 3. Results of summaries with ROUGE-1. ***(The higher the value, the higher the quality of the summarisation result)***

<table>
<tr><td>Sample 1</td><td colspan="3">Refer to Appendix, Sample 1.</td></tr>
<tr><td>TextRank</td><td colspan="3">The shaking lasted about 50 seconds, said CNN meteorologist Chad Myers. SAN FRANCISCO, California (CNN) -- A magnitude 4.2 earthquake shook the San Francisco area Friday at 4:42 a.m. PT, the U.S. Geological Survey reported. "We had quite a spike in calls, mostly calls of inquiry, none of any injury, none of any damage that was reported," said Capt. According to the USGS, magnitude 4.2 quakes are felt indoors and may break dishes and windows and overturn unstable objects. The quake left about 2000 customers without power, said David Eisenhower, a spokesman for Pacific Gas and Light.</td></tr>
<tr><td>LongT5Rank</td><td colspan="3">The earthquake lasted about fifty seconds and was felt throughout the area. No one was seriously injured. A spokesman from Pacific Gas and light says that it was not a big quake but rather a minor one.</td></tr>
<tr><td rowspan="4">ROUGE-1</td><td></td><td>TextRank</td><td>LongT5Rank</td></tr>
<tr><td>Precision</td><td>0.28260869 ✓</td><td>0.15217391</td></tr>
<tr><td>Recall</td><td>0.12871287</td><td>0.18918918 ✓</td></tr>
<tr><td>F-Measure</td><td>0.17687074 ✓</td><td>0.16867469</td></tr>
<tr><td>Sample 2</td><td colspan="3">Refer to Appendix, Sample 2.</td></tr>
<tr><td>TextRank</td><td colspan="3">There is a properly political sphere in which the monarch may counsel but doesn't intrude, but there is another sphere, vital to any sense of national unity and creativity, a sphere in which communities must be celebrated, common values articulated and the transcendent source of those values honored. Now to Him who by the power at work within us is able to accomplish abundantly far more than all we can ask or imagine, to Him be glory in the church and in Christ Jesus to all generations, for ever and ever. It must have been a bewildering time for the Princess as well, but even then, at the age of 20, her capacity for empathy and her very strong intuitive power ensured that any tension soon evaporated. Let this service mark the point at which we let her rest in peace and dwell on her memory with thanksgiving and compassion. The role brought great power but, like any member of the Royal Family, she also experienced the weight of expectation and the intensity of the scrutiny. She said of John, a young Greek suffering from cystic fibrosis: "He showed no sign of anger, no trace of bitterness but touched us all with an aura of optimism and hope for the future such that I have never before encountered." She sought out places of suffering, because they are so very often places of truth where the masks have been removed, and she was not afraid to be with the dying and to comfort them in an unsentimental way. LONDON, England (CNN) -- Prince Harry led tributes to Diana, Princess of Wales on the 10th anniversary of her death, describing her as "the best mother in the world" in a speech at a memorial service. Those familiar with the field have no doubt that the Princess played a significant part in overcoming a harmful and even a cruel taboo in a gesture which was not choreographed but sprung from a deep identification with those who were vulnerable and on the margin. Led by our Queen and other members of the Royal Family, our constitution has developed in response to the challenges of the past century. Princess Diana recognized this quality of life in many of those, like John, whose lives she touched. Princes Harry and William greet guests at a thanksgiving service in memory of their mother. But behind the media glare, to us, just two loving children, she was quite simply the best mother in the world. To lose a parent so suddenly at such a young age, as others have experienced, is indescribably shocking and sad. An editorial in the New York Times in 1989 admitted ruefully that it had taken a foreign, and even a royal, dignitary to draw attention to a major public health concern in the US. I pray that you may have the power to comprehend, with all the saints, what is the breadth</td></tr>
</table>

https://doi.org/10.18080/jtde.v12n3.977

<table>
<tr><td></td><td colspan="3">and length and height and depth, and to know the love of Christ that surpasses knowledge, so that you may be filled with all the fullness of God. Still, 10 years after her tragic death, there are regular reports of "fury" at this or that incident, and the Princess's memory is used for scoring points. And as she said, in her words, "the biggest disease today is not leprosy or TB but the feeling of being unwanted". "I pray that you being rooted and established in love may have power with all the saints to grasp what is the breadth and length and depth and height of the love of Christ and to know this love which surpasses knowledge that you might be filled with the fullness of God." Honoring but managing the role and not allowing it to take over one's personal humanity is a desperately difficult task. This is certain and has been proved experimentally in the life of all the saints. Let us also, echoing the words of Prince Harry, look to the future and pray, in the words of St Paul, for all those who serve our country as members of the Royal Family and most especially for the sons who were so precious to her: . The love of Christ described in the lesson read by Prince William contains the essence of the spiritual life. I pray that, according to the riches of His glory, He may grant that you may be strengthened in your inner being with power through His Spirit, and that Christ may dwell in your hearts through faith, as you are being rooted and grounded in love. Afterwards they comment on her large eyes and what life she brought into the room. Her beaming smile greeted us from school. She -- like our father -- was determined to provide us with a stable and secure childhood. We tend to be suspicious of public figures who wrap themselves in divinity and claim that their will is God's will, but if no-one can articulate in an un-ignorable way in the public realm the creative energy of the love that we see in Christ, the human face of God, then we shall find ourselves inhabiting a maimed and diminished society. Two residents are playing Beggar My Neighbor. Her work in the very last year of her life for the victims of landmines also caught the popular imagination internationally and certainly accelerated the adoption of the Ottawa Convention, banning the use of a weapon which disproportionately kills and maims women and children. She confessed to receiving a very great deal from some of those whose lives she touched. It's easy to lose the real person in the image, to insist that all is darkness or all is light. The question from the royal visitor is unexpected but everyone laughs.</td></tr>
<tr><td>LongT5Rank</td><td colspan="3">In this chapter, the narrator gives us a little background on Diana's life and then proceeds to give us some reflections about how important it is for the royal family to be able to share in the good work that God is doing in the world through his Son. We learn that Diana was very kind to her children when they were young, and that she also played a big role in making them feel better after their mother's death. Prince Harry and William are particularly fond of Diana because she was such a great mom. They ask her to help them understand the depth and breadth of Christ's love so that they can have faith in him.</td></tr>
<tr><td rowspan="4">ROUGE-1</td><td></td><td>TextRank</td><td>LongT5Rank</td></tr>
<tr><td>Precision</td><td>0.6666666 ✓</td><td>0.4222222</td></tr>
<tr><td>Recall</td><td>0.0279792</td><td>0.1570247 ✓</td></tr>
<tr><td>F-Measure</td><td>0.0534653</td><td>0.2289156 ✓</td></tr>
<tr><td>Sample 3</td><td colspan="3">Refer to Appendix, Sample 3.</td></tr>
<tr><td>TextRank</td><td colspan="3">Spain's Foreign Ministry has publicly expressed its disagreement with the charges and has dispatched top diplomats to Chad to try to win the group's release. The Spanish flight crew is innocent and should be released, a company executive said Thursday. Other French charities earlier had questioned whether Zoe's Ark could legally arrange adoption of children from Darfur, and contacted French authorities, according to French newspapers and The Associated Press. "So far, the interviews carried out with the children -- some of whom could not provide any information due to their young age -- led to the preliminary conclusion that probably 85 come from Chadian villages near the cities of Adre and Tine along the Chadian-Sudanese border," the agencies said. French authorities have reacted angrily to the Zoe's Ark trip, calling the group's actions "illegal and irresponsible." The charity workers and journalists have been charged with kidnapping and extortion and could face 20 years of hard labor if convicted. The third journalist, Marie-Agnes Peleran of the TV station France 3 Miditerranee, was traveling with the group in a personal capacity, though she carried a camera from her station, Reporters Without Borders said. The French Foreign Ministry has said the dispute will not affect France's participation in a European peacekeeping force due to be deployed along the border between Chad and Sudan. ABECHE, Chad (CNN) -- Most of the 103 children that a French charity attempted to take to France from Chad for adoption are neither Sudanese nor orphans, three international aid agencies reported on Thursday. Six members of Zoe's Ark were arrested last week as they tried to put the children on a</td></tr>
</table>

	plane to France, where the charity said host families were waiting to take the children in. Chadian President Idriss Deby hopes the journalists and the flight crew will be freed, his chief of staff, Mahamat Hissene, said Thursday. On its Web site, Zoe's Ark describes itself as a nonprofit organization based in Paris that sends teams of physicians, nurses, firefighters and other specialists to care for children in war zones and place them with families in France, who then apply for asylum on their behalf. In response to the dispute in Chad, the Republic of Congo said late Wednesday it was suspending all international adoptions, The Associated Press reported. The charity said the children were Sudanese orphans that it was trying to rescue from a war-torn nation. The president would legally be able to intervene in the case if it is transferred from a judge in the eastern city of Abeche, where the children were taken, to a judge in N'Djamena, the capital, Hissene said. "Ninety-one children said they had been living with their family, consisting of at least one adult they considered to be their parent," the agencies said, adding that interviews with the remaining 12 children were ongoing.		
LongT5Rank	Spain's foreign ministry says it disagrees with the charges, and dispatches top diplomatics to try to free the group. The Spanish Flight Crew is innocent but should be released. Some French charities had earlier asked whether Zoe's Ark can legally arrange adoption for children from Darhur, and consulted French authorities. The agencies said that some of the children came from Chadian village near the cities Adre and Tie along the border. French authorities are furious at the group because they have been accused of "kidnapping" and "extortion." Three journalists were traveling with them in their personal capacity. The French Ministry has said its dispute will not effect France's involvement in a Europe-based peacekeeping force which is due to be deployed between Chad and and Sudan.		
ROUGE-1		**TextRank**	**LongT5Rank**
	Precision	0.83333333 ✓	0.30555555
	Recall	0.19230769	0.33333333 ✓
	F-Measure	0.3125	**0.31884057** ✓

Human-level performance

In text summarisation, human-level performance refers to a machine-learning model's capacity to create summaries that are as excellent as or indistinguishable from summaries written by people. This includes not just catching the essential points of the original text, but also preserving coherence, readability, and grammatical accuracy (Vodolazova & Lloret, 2019). Hence, it can be used to evaluate the quality of the summaries generated by the proposed system.

Semantic Textual Similarity and ROUGE provide a quantitative way to evaluate the performance of the proposed system. However, these metrics may not always align with human judgment. This is because those machines, unlike humans, do not truly understand text in the way humans do. Therefore, Human Level Performance (HLP) should be used in the evaluation stage.

HLP helps in examining whether the system can be able to perform a task that requires human understanding. A human-level Natural Language Processing should be able to provide an accurate and trustworthy summary that can be interpreted by humans in order for them to understand and get reliable information.

Human evaluation has been taken via informed reviews by asking the users to rate the summary output generated by the LongT5Rank summariser. The criteria in evaluating the

https://doi.org/10.18080/jtde.v12n3.977

summary quality will be in terms of accuracy, relevance and coherence. By having this human evaluation, people can provide instant and direct feedback on how well the text summariser is meeting their expectations and needs.

From the pie charts in Figure 3 and Figure 4, it can be seen that the combined TextRank and LongT5 method has a larger percentage of "Accurate & Coherent" summaries compared to TextRank alone. This is due to the fact that the TextRank algorithm just pulls the most important sentences directly from the source text without considering the coherence of the sentences. Hence, the summary generated by the TextRank algorithm cannot easily be understood by humans. This suggests that the combined method is more effective at generating a summary that is both accurate in content and yet coherent in structure, so that it can be understood by humans. Not only that, users also gave the feedback about the generated summaries that they help to reduce the reading time by half compared to reading the entire lengthy texts.

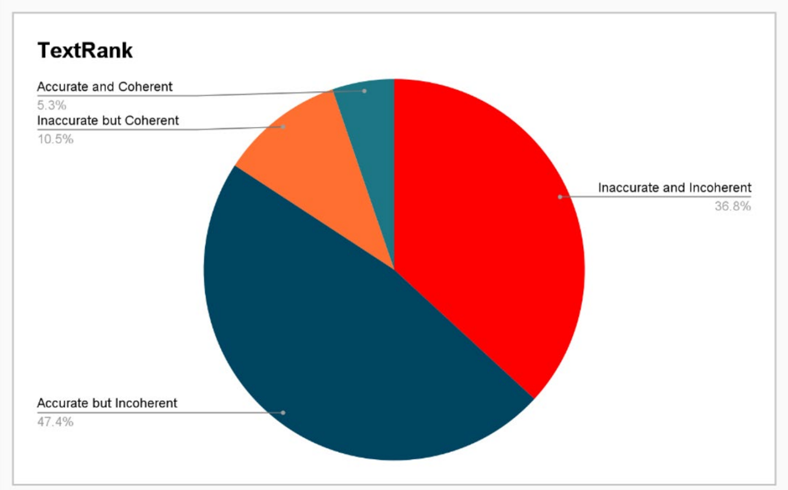


Figure 3. Results of TextRank summaries with HLP evaluation

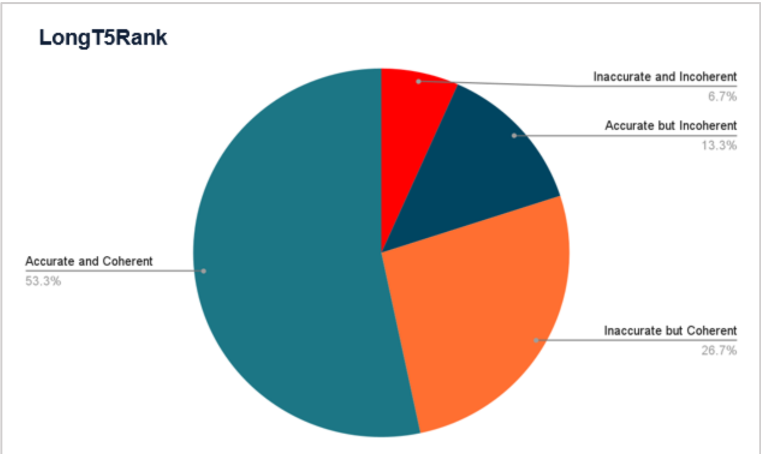


Figure 4. Result of LongT5Rank summaries with HLP evaluation

Compression

The compression rate refers to the amount of text length that is reduced during the summarisation process. It controls how much information is maintained in the summary versus the original text. The formula to calculate the compression rate is as follows:

$$\textit{Compression Rate} = \frac{\textit{Original Length - Summary Length}}{\textit{Original Length}} \tag{5}$$

The "Summary Length Ratio" represents the ratio of the summarized text length to the original text length. The "Compression Rate" shows the proportion of the original text that remains after summarization, indicating how concisely the article was summarized. The results in Table 4 demonstrate that the proposed system consistently achieves a compression rate above 60%, ensuring that the lengthy article is concisely summarized while retaining the essential information.

Table 4. Results of Summaries with Compression Rate as a Measure

Summary Length Ratio	Compression Rate
0.2	0.827
0.4	0.799
0.6	0.766
0.8	0.614

Benchmarking with other methods

To evaluate the performance of LongT5Rank, we compared it with other well-known summary methods, including TextRank, BERTSUM (Torres, 2021), and BART (Lewis et al., 2019). We employed Precision, Recall and F-measure as comparative measures. Table 5 shows the results of these benchmarks with Sample 2.

From the results, TextRank has the greatest Precision (0.66666) of all the methods, suggesting that it picks phrases that are extremely relevant to the original text. However, it has the lowest Recall (0.0279), indicating that it misses a significant amount of crucial information. As a result, its F-measure (0.0534) is equally low.

BERTSUM outperforms TextRank on a balance of Recall (0.0781) and F-measure (0.1369). This shows that it catches more relevant information, although with a slightly lower Precision (0.5555) than TextRank.

BART has the greatest Recall (0.2916) of the approaches, indicating that it contains a large quantity of meaningful information in its summaries. It has a lesser Precision (0.3111) than TextRank and BERTSUM, but the best overall F-measure (0.3010), indicating a strong mix of precision and recall.

Our suggested LongT5Rank technique has a balanced performance, with a Precision of 0.42222, Recall of 0.1570, and F-measure of 0.2289. While it may not obtain the best score in any single criterion, it performs well across all parameters, demonstrating its ability to provide short and cohesive summaries.

Table 5. Performance Comparison of Summarisation Methods

Method	Precision	Recall	F-Measure
TextRank	0.66666	0.0279	0.0534
BERTSUM	0.5555	0.0781	0.1369
BART	0.3111	0.2916	0.3010
LongT5Rank	0.42222	0.1570	0.2289

Conclusion

The integration, LongT5Rank, of TextRank with the LongT5 model in the text summarisation system efficiently addresses the challenge of extracting information from lengthy texts. The system's achievements include a minimum compression rate of 60%, a semantic textual similarity score of 0.6, and an increased F-measure score for precision and recall. User customization options, such as input types and summary length, enhance flexibility. Overall, this innovative and adaptable system represents a significant advance in natural language processing, offering a user-centric solution to the complexities of information retrieval.

Acknowledgements

The authors would like to express their gratitude to the faculty and staff of Tunku Abdul Rahman University of Management and Technology (TAR UMT) for their invaluable support and guidance throughout the research project. This research was made possible with the assistance of various libraries, online resources, and funding from TAR UMT. Finally, we extend our heartfelt appreciation to all those who participated in the evaluation and testing phases of the project, without whom this study would not have been possible.

References

Abujar, S., Masum, A. K. M., Sanzidul Islam, M., Faisal, F., & Hossain, S. A. (2020). A Bengali Text Generation Approach in Context of Abstractive Text Summarization Using RNN. In Saini, H., Sayal, R., Buyya, R., & Aliseri, G. (eds), *Innovations in Computer Science and Engineering*. *Lecture Notes in Networks and Systems*, vol. 103. Springer, Singapore. pp. 509–518. https://doi.org/10.1007/978-981-15-2043-3_55

Bichi, A. A., Samsudin, R., Hassan, R., Hasan, L. R. A., & Ado Rogo, A. (2023). Graph-based extractive text summarization method for Hausa text. *PloS One*, *18*(5), e0285376–e0285376. https://doi.org/10.1371/journal.pone.0285376

Chawla, J. S. (2020, July 6). Word Vectorization using GloVe. Analytics Vidhya. https://medium.com/analytics-vidhya/word-vectorization-using-glove-76919685ee0b

Dewda, M. (2022, September 18). Abstractive Text Summarization. Globant. https://medium.com/globant/abstractive-text-summarization-bccb4bf5851c

El-Kassas, W. S., Salama, C. R., Rafea, A. A., & Mohamed, H. K. (2020). Automatic Text Summarization: A Comprehensive Survey. *Expert Systems with Applications, 165*, 113679. https://doi.org/10.1016/j.eswa.2020.113679

Gambhir, M., & Gupta, V. (2016). Recent automatic text summarization techniques: a survey. *Artificial Intelligence Review, 47*(1), 1–66. https://doi.org/10.1007/s10462-016-9475-9

Ghadimi, A., & Beigy, H. (2022). Hybrid multi-document summarization using pre-trained language models. *Expert Systems with Applications, 192*, 116292. https://doi.org/10.1016/j.eswa.2021.116292

Guo, M., Ainslie, J., Uthus, D., Ontanon, S., Ni, J., Sung, Y.-H., & Yang, Y. (2022). LongT5: Efficient Text-To-Text Transformer for Long Sequences. ArXiv:2112.07916 [Cs]. https://arxiv.org/abs/2112.07916

Hernandez-Castaneda, A., Garcia-Hernandez, R. A., Ledeneva, Y., & Millan-Hernandez, C. E. (2020). Extractive Automatic Text Summarization Based on Lexical-Semantic Keywords. *IEEE Access, 8*, 49896–49907. https://doi.org/10.1109/access.2020.2980226

Jayan, J. P., & Govindaru, G. (2022). Automatic Summarization of Malayalam Documents using Text Extraction Methods. SCRS Conference Proceedings on Intelligent Systems, 443–457. https://www.publications.scrs.in/chapter/978-93-91842-08-6/42

Joshi, P. (2023, May 22). An Introduction to Text Summarization using the TextRank Algorithm (with Python implementation). Analytics Vidhya. https://www.analyticsvidhya.com/blog/2018/11/introduction-text-summarization-textrank-python/

Khor, Y. K., Tan, C. L., & Lim, T. M. (2022). Extractive Summarization on Food Reviews. *Journal of The Institution of Engineers, Malaysia—ICDXA Special Issue, 82*(3). https://doi.org/10.54552/v82i3.96

Krishnan, S. (2022, January 5). Why Cosine Similarity is used in Natural Language Processing? Medium. https://sandhyakrishnan02.medium.com/cosine-similarity-for-natural-language-processing-d761e2c02d10

Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., & Zettlemoyer, L. (2019). BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. ArXiv:1910.13461 [Cs, Stat]. https://arxiv.org/abs/1910.13461

Luo, C., Chen, Z., Jiang, X., & Yang, S. (2022). Gap Sentences Generation with TextRank for Chinese Text Summarization. ACAI '22: Proceedings of the 2022 5th International Conference on Algorithms, Computing and Artificial Intelligence. Article 67, pp 1–5. https://doi.org/10.1145/3579654.3579725

Majumder, G., Pakray, P., Gelbukh, A., & Pinto, D. (2016). Semantic Textual Similarity Methods, Tools, and Applications: A Survey. *Computación y Sistemas, 20*(4). https://doi.org/10.13053/cys-20-4-2506

Mishra, U. (2022). What Is Text Summarization in NLP? Analytics Steps. www.analyticssteps.com. https://www.analyticssteps.com/blogs/what-text-summarization-nlp

Priyanka. (2022, November 21). ROUGE your NLP Results! Medium. https://medium.com/@priyankads/rouge-your-nlp-results-b2feba61053a

Rusnachenko, N. (2024, January 5). nicolay-r/ViLongT5. GitHub. https://github.com/nicolay-r/ViLongT5

Santhosh, S. (2023, April 16). Understanding BLEU and ROUGE score for NLP evaluation. Medium. https://medium.com/@sthanikamsanthosh1994/understanding-bleu-and-rouge-score-for-nlp-evaluation-1ab334ecadcb

Torres, S. (2021). Evaluating Extractive Text Summarization with BERTSUM Stanford CS224N Custom Project. Retrieved August 25, 2024, from https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1214/reports/final_reports/report042.pdf

Vodolazova, T., & Lloret, E. (2019, September 1). Towards Adaptive Text Summarization: How Does Compression Rate Affect Summary Readability of L2 Texts? In Mitkov, R. & Angelova, G. (eds.). Proceedings of Recent Advances in Natural Language Processing, pp. 1265–1274. https://doi.org/10.26615/978-954-452-056-4_145

Zaware, S., Patadiya, D., Gaikwad, A., Gulhane, S., & Thakare, A. (2021). Text Summarization using TF-IDF and Textrank algorithm. 2021 5th International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2021, pp. 1399–1407. https://doi.org/10.1109/ICOEI51242.2021.9453071

Appendix

Sample 1

SAN FRANCISCO, California (CNN) -- A magnitude 4.2 earthquake shook the San Francisco area Friday at 4:42 a.m. PT, the U.S. Geological Survey reported. The quake left about 2000 customers without power, said David Eisenhower, a spokesman for Pacific Gas and Light. Under the USGS classification, a magnitude 4.2 earthquake is considered "light," which it says usually causes minimal damage. "We had quite a spike in calls, mostly calls of inquiry, none of any injury, none of any damage that was reported," said Capt. Al Casciato of the San Francisco police. "It was fairly mild." Watch police describe concerned calls immediately after the quake. The quake was centered about two miles east-northeast of Oakland, at a depth of 3.6 miles, the USGS said. Oakland is just east of San Francisco, across San Francisco Bay. An Oakland police dispatcher told CNN the quake set off alarms at people's homes. The shaking lasted about 50 seconds, said CNN meteorologist Chad Myers. According to the USGS, magnitude 4.2 quakes are felt indoors and may break dishes and windows and overturn unstable objects. Pendulum clocks may stop.

Sample 2

LONDON, England (CNN) -- Prince Harry led tributes to Diana, Princess of Wales on the 10th anniversary of her death, describing her as "the best mother in the world" in a speech at a memorial service. Here is his speech in full: William and I can separate life into two parts. There were those years when we were blessed with the physical presence beside us of both our mother and father. Princes Harry and William greet guests at a thanksgiving service in memory of their mother. And then there are the 10 years since our mother's death. When she was alive, we completely took for granted her unrivaled love of life, laughter, fun and folly. She was our guardian, friend and protector. She never once allowed her unfaltering love for us to go unspoken or undemonstrated. She will always be remembered for her amazing public work. But behind the media glare, to us, just two loving children, she was quite simply the best mother in the world. We would say that, wouldn't we. But we miss her. She kissed us last thing at night. Her beaming smile greeted us from school. She laughed hysterically and uncontrollably when sharing something silly she might have said or done that day. She encouraged us when we were nervous or unsure. She -- like our father -- was determined to provide us with a stable and secure childhood. To lose a parent so suddenly at such a young age, as others have experienced, is indescribably shocking and sad. It was an event which changed our lives forever, as it must have done for everyone who lost someone that night. But what is far more important to us now, and into the future, is that we remember our mother as she would have wished to be remembered as she was: fun-loving, generous, down-to-earth, entirely genuine. We both think of her every day. We speak about her and laugh together at all the memories. Put simply, she made us, and so many other people, happy. May this be the way that she is remembered. Prince William's reading from St Paul's letter to the Ephesians: . I bow my knees before the Father, from whom every family in heaven and on earth takes its name. I pray that, according to the riches of His glory, He may grant that you may be strengthened in your inner being with power through His Spirit, and that Christ may dwell in your hearts through faith, as you are being rooted and grounded in love. I pray that you may have the power to comprehend, with all the saints, what is the breadth and length and height and depth, and to know the love of Christ that surpasses knowledge, so that you may be filled with all the fullness of God. Now to Him who by the power at work within us is able to accomplish abundantly far more than all we can ask or imagine, to Him be glory in the church and in Christ Jesus to all generations, for ever and ever. Amen. Thanks be to God. The Bishop of London's speech: . "Who's cheating?" The scene is an old people's home. Two residents are playing Beggar My Neighbor. Enter the Princess. The question from the royal visitor is unexpected but everyone laughs. Afterwards they comment on her large eyes and what life she brought into the room. One tiny incident, characteristic of countless other occasions in the Princess's public life in which she found the right word or the right gesture to bring cheer and comfort. Everyone here will have their own memories. I remember meeting Princess Diana for the very first time early in 1981 to discuss details of the wedding service in St Paul's. Even Archbishop's Chaplains have their share of proper diffidence and I was nervous entering the presence. It must have been a bewildering time for the Princess as well, but even then, at the age of 20, her capacity for empathy and her very strong intuitive power ensured that any tension soon evaporated. Prince Harry has spoken movingly and justly, as few others have the right to do, about the Princess as a mother. I want to dwell for a moment on her public work, its cost and its meaning. After her marriage, the Princess joined her natural gifts of beauty, empathy and powerful intuition with that extraordinary charge which association with the Royal Family generates. Led by our Queen and other members of the Royal Family, our constitution has developed in response to the challenges of the past century. There is a properly political sphere in which the monarch may counsel but doesn't intrude, but there is another sphere, vital to any sense of national unity and creativity, a sphere in which communities must be celebrated, common values articulated and the transcendent source of those values honored. We tend to be suspicious of public figures who wrap themselves in divinity and claim that their will is God's will, but if no-one can articulate in an un-ignorable way in the public realm the creative energy of the love that we see in Christ, the human face of God, then we shall find ourselves inhabiting a maimed and diminished society. And at a time when people are suspicious of rhetoric, the monarchy communicates by symbol and by simple speech, and the Princess brought her own gifts to this work. She was still only 26 in 1987 when she shook the hand of a patient at the opening of the Middlesex Hospital's Aids ward. It was the first in the UK and it is very hard now to credit the degree of fear and prejudice which surrounded Aids in the '80s. Those familiar with the field have no doubt that the Princess played a significant part in overcoming a harmful and even a cruel taboo in a gesture which was not choreographed but sprung from a deep identification with those who were vulnerable and on the margin. And she had a similar impact in the USA. An editorial in the New York Times in 1989 admitted ruefully that it had taken a foreign, and even a royal, dignitary to draw attention to a major public health concern in the US. Her work in the very last year of her life for the victims of landmines also caught the popular imagination internationally and certainly accelerated the adoption of the Ottawa Convention, banning the use of a weapon which disproportionately kills and maims women and children. She proved the eloquence of embrace and of touch which, of course, have been used by royal healers throughout the centuries. And as she said, in her words, "the biggest disease today is not leprosy or TB but the feeling of being unwanted". She sought out places of suffering, because they are so very often places of truth where the masks have been removed, and she was not afraid to be with the dying and to comfort them in an unsentimental way. Bill Deedes accompanied her on some of her visits. His response to the cynics was typically robust. He said: "She was one who sought above all to help vulnerable people in society and who did it so well. She was good at this because she herself was vulnerable. She knew the feeling. She didn't set out to be a saint." The role brought great power but, like any member of the Royal Family, she also experienced the weight of expectation and the intensity of the scrutiny. Honoring but managing the role and not allowing it to take over one's personal humanity is a desperately difficult task. As we have heard from Prince Harry, his mother Diana did all that she could to prepare her sons for the work which lies ahead. She confessed to receiving a very great deal from some of those whose lives she touched. She said of John, a young Greek suffering from cystic fibrosis: "He showed no sign of anger, no trace of bitterness but touched us all with an aura of optimism and hope for the future such that I have never before encountered." The love of Christ described in the lesson read by Prince William contains the essence of the spiritual life. Princess Diana recognized this quality of life in many of those, like John, whose lives she touched. It was a mystery which resonated deeply with her and for which she reached out. And the mystery is this - the more you go beyond yourself, the more you will become your true self; the more you lose yourself in loving and serving others, the more you will find yourself; the more you keep company with those who suffer, the more you will be healed. This is the knowledge which passes all understanding. This is certain and has been proved experimentally in the life of all the saints. It's easy to lose the real person in the image, to insist that all is darkness or all is light. Still, 10 years after her tragic death, there are regular reports of "fury" at this or that incident, and the Princess's memory is used for scoring points. Let it end here. Let this service mark the point at which we let her rest in peace and dwell on her memory with thanksgiving and compassion. Let us also, echoing the words of Prince Harry, look to the future and pray, in the words of St Paul, for all those who serve our country as members of the Royal Family and most especially for the sons who were so precious to her: . "I pray that you being rooted and established in love may have power with all the saints to grasp what is the breadth and length and depth and height of the love of Christ and to know this love which surpasses knowledge that you might be filled with the fullness of God." Amen.

Sample 3

ABECHE, Chad (CNN) -- Most of the 103 children that a French charity attempted to take to France from Chad for adoption are neither Sudanese nor orphans, three international aid agencies reported on Thursday. Hundreds of women protest child trafficking and shout anti-French slogans Wednesday in Abeche, Chad. Six members of Zoe's Ark were arrested last week as they tried to put the children on a plane to France, where the charity said host families were waiting to take the children in. Three French journalists, a seven-member Spanish flight crew and one Belgian were also arrested. Representatives of the journalists and flight crew said they were unaware of problems with Zoe's Ark and thought they were on a humanitarian mission. Chadian President Idriss Deby hopes the journalists and the flight crew will be freed, his chief of staff, Mahamat Hissene, said Thursday. The president would legally be able to intervene in the case if it is transferred from a judge in the eastern city of Abeche, where the children were taken, to a judge in N'Djamena, the capital, Hissene said. The transfer will take place Monday, according to media reports. The International Red Cross Committee, the U.N. High Commissioner for Refugees and UNICEF said most of the children were living with their families before Zoe's Ark took them. The charity said the children were Sudanese orphans that it was trying to rescue from a war-torn nation. The agencies said most of the children also probably come from Chadian villages along Chad's border with Sudan. The children have been living in an orphanage in Abeche while authorities and aid agencies try to determine their identities. Watch a report on whether the children are orphans » . Chadian authorities immediately accused the charity of kidnapping the children and concealing their identities. Chad's interior minister said Zoe's Ark dressed the children in bandages and fake intravenous drips to make them look like refugees who needed medical help. The charity workers and journalists have been charged with kidnapping and extortion and could face 20 years of hard labor if convicted. The Spaniards and Belgian are charged with complicity. The Spanish flight crew is innocent and should be released, a company executive said Thursday. "We thought we were doing a humanitarian transport," said Antoni Cajal, sales director of Spain's Gir Jet charter firm. "If an NGO [nongovernmental organization] has done something wrong, it's impossible for us to know." Spain's Foreign Ministry has publicly expressed its disagreement with the charges and has dispatched top diplomats to Chad to try to win the group's release. Over the weekend, the captain appealed urgently to be rescued, fearing the crew could be harmed or killed, Cajal said. But the four women and three men are in good condition in custody, Cajal said, based on his conversations with a Spanish consular official who came from Cameroon to Chad and has been able to visit them. The detention is the first problem of its kind for the company, which hopes government negotiations can resolve the issue, Cajal said. On its Web site, Zoe's Ark describes itself as a nonprofit organization based in Paris that sends teams of physicians, nurses, firefighters and other specialists to care for children in war zones and place them with families in France, who then apply for asylum on their behalf. The Red Cross, UNHCR and UNICEF said the 21 girls and 82 boys range in age from about 1 year to about 10, and they are healthy. The agencies said they have been interviewing the children individually to determine their backgrounds. "So far, the interviews carried out with the children -- some of whom could not provide any information due to their young age -- led to the preliminary conclusion that probably 85 come from Chadian villages near the cities of Adre and Tine along the Chadian-Sudanese border," the agencies said. "Ninety-one children said they had been living with their family, consisting of at least one adult they considered to be their parent," the agencies said, adding that interviews with the remaining 12 children were ongoing. The agencies called their investigation painstaking and challenging because of the number of children, their youth and the situation in the region. Other French charities earlier had questioned whether Zoe's Ark could legally arrange adoption of children from Darfur, and contacted French authorities, according to French newspapers and The Associated Press. French authorities have reacted angrily to the Zoe's Ark trip, calling the group's actions "illegal and irresponsible." The French Foreign Ministry has said the dispute will not affect France's participation in a European peacekeeping force due to be deployed along the border between Chad and Sudan. In response to the dispute in Chad, the Republic of Congo said late Wednesday it was suspending all international adoptions, The Associated Press reported. Reporters Without Borders said it will work for the release of the three journalists arrested in Chad. The organization said photographers Marc Garmirian of the Capa news agency and Jean-Daniel Guillou of the Synchro X agency were on assignment for their news organizations and were not part of the charity's efforts. The third journalist, Marie-Agnes Peleran of the TV station France 3 Miditerranee, was traveling with the group in a personal capacity, though she carried a camera from her station, Reporters Without Borders said. E-mail to a friend. CNN's Al Goodman contributed to this report. Copyright 2007 CNN. All rights reserved. This material may not be published, broadcast, rewritten, or redistributed.

Exploring Customer Segmentation in E-Commerce using RFM Analysis with Clustering Techniques

Chun-Gee Wong
Multimedia University, Malaysia

Gee-Kok Tong
Multimedia University, Malaysia

Su-Cheng Haw
Multimedia University, Malaysia

Abstract: The proliferation of big data and the growth of e-commerce have intensified the challenges associated with extracting actionable data for personalised recommendations and decision-making. With data-driven marketing strategies, understanding and predicting customer behaviour has become paramount for maintaining competitive advantage. This study leverages business analytics tools, focusing on Recency, Frequency, and Monetary (RFM) Analysis, alongside K-Means and Hierarchical (Agglomerative) Clustering algorithms, to segment customer transactional data. Data normalisation, a critical step for accurate clustering, was performed using log transformation and the Power Transformer technique with the Yeo-Johnson parameter, the latter proving more effective for handling both positively and negatively skewed data, enhancing data normalisation and suitability for analysis. This study reveals that RFM Analysis with Hierarchical Clustering outperforms K-Means Clustering, achieving a Silhouette Score of 0.47 and a Calinski–Harabasz Index of 3787.1, indicating a more accurate identification of customer segments. RFM Analysis alone generated eight clusters, while integrating RFM Analysis with both Hierarchical Clustering and K-Means generated three similar-sized clusters with interchanged labels. These metrics highlight the proficiency of Hierarchical Clustering in identifying unique customer segments and customising marketing strategies. The findings indicate that the RFM-Hierarchical Clustering approach enhances segmentation precision and facilitates more refined and effective marketing strategies.

Keywords: customer segmentation, RFM Analysis, Clustering Analysis, K-Means Clustering, Hierarchical Clustering

https://doi.org/10.18080/jtde.v12n3.978

Introduction

The importance of big data in e-commerce has grown significantly in contemporary societies, necessitating the adoption of business analytics (BA) tools to inform marketing decisions (Li, 2021; Raguseo & Vitari, 2018). Mallam *et al.* (2021) affirm that data-driven decision-making has not only increased work productivity but has also helped businesses to achieve long-term objectives. Additionally, the COVID-19 pandemic has further accelerated the expansion and demand for e-commerce as industries seek to leverage virtual spaces to transform traditional in-person trading within markets and organisations (Garg *et al.*, 2021).

Fahrudin & Rindiyani (2024) highlight that while many established online retail brands have successfully leveraged data mining to achieve a competitive advantage, smaller and newer retailers frequently face challenges in adopting these technologies. They further assert that the use of data mining and clustering techniques is crucial for identifying customer segments through detailed RFM characteristics and patterns, which can be hard to detect with only manual analysis. This difficulty often stems from a lack of the necessary technical knowledge and expertise required to effectively implement consumer-centric marketing strategies (Chen *et al.*, 2012). In this competitive business environment, rivalries among organisations have intensified, making it crucial to attract and retain customers. Consequently, delivering exceptional customer service has become critical for businesses of all sizes (Sarkar *et al.*, 2024).

Despite these advancements, Kaur *et al.* (2022) observe that e-commerce businesses still face significant challenges in effectively identifying and utilising data for decision-making and personalisation within the realm of business intelligence. One strategy to address these challenges is customer segmentation, a technique widely used in the retail industry to fine-tune marketing initiatives by catering to specific customer groups, thus more effectively meeting their individual needs and preferences (Idowu & Kattukottai, 2019).

Understanding consumer behaviour is another critical aspect for businesses aiming to thrive. Mehmeti & Luga (2021) discuss the complexity of consumer behaviour, which is influenced by various factors and crucial for developing effective marketing strategies to enhance sales revenue and customer retention. Additionally, Serwah *et al.* (2023) emphasise the necessity for online retailers to continually adapt to fierce competition and rapidly changing market conditions by employing customer segmentation to organise customers into smaller, more uniform groups based on shared characteristics.

https://doi.org/10.18080/jtde.v12n3.978

Problem Statement

Despite advancements in e-commerce analytics, a significant gap remains in optimising customer segmentation methodologies for enhanced personalisation. This gap highlights the need for generating new knowledge on effective tools and techniques to analyse and segment customer data accurately.

Research questions

1. How do Hierarchical Clustering and K-Means Clustering compare in their effectiveness for customer segmentation in e-commerce?
2. What are the impacts of integrating RFM (Recency, Frequency and Monetary) Analysis with clustering techniques on the accuracy of customer segmentation?
3. How does the application of different data normalisation techniques (log transformation versus Power Transformer with Yeo-Johnson parameter) affect the accuracy and quality of customer segmentation in e-commerce?

Objectives

This study aims to:

1. Analyse the effectiveness of RFM Analysis in customer segmentation.
2. Compare the performance of Hierarchical and K-Means Clustering in segmenting e-commerce customers, focusing on metrics such as the Silhouette Score and Calinski–Harabasz Index.
3. Provide actionable insights for personalised marketing strategies based on the findings.

This research utilises BA tools, specifically RFM Analysis, along with clustering techniques such as K-Means and Hierarchical Clustering, to increase the understanding of customer behaviour and improve customer segmentation efforts. This methodological approach enables businesses to use RFM Analysis, K-Means and Hierarchical Clustering to develop more effective marketing strategies and gain comprehensive insights into customer dynamics.

Related Works

This section offers a concise review of prior studies relevant to our topic, highlighting significant contributions and findings from other researchers.

Customer segmentation and customer behaviour

Customer segmentation entails dividing customers into distinct groups according to various characteristics, including behavioural, social and consumption patterns, to tailor different marketing strategies (Kadir & Achyar, 2019). Kurniawan *et al.* (2018) note that consumer behaviour encompasses the actions and decision-making processes of consumers regarding the purchase and use of goods and services. Additionally, individual demands and interests, as well as distinct behavioural traits, significantly influence these decisions. Despite these differences, customers share a common desire to achieve maximum satisfaction in their consumption.

Moreover, previous research has demonstrated that customers can be grouped according to variables such as past purchases and visit frequency (Anitha & Patil, 2022; Griva *et al.*, 2018). For customer segmentation to be truly effective and beneficial, the collected information must be comprehensible, accessible, relevant, and significant (Serwah *et al.*, 2023; Anitha & Patil, 2022; Chaubey *et al.*, 2022; Idowu & Kattukottai, 2019). A variety of methods for segmenting customers and analysing their purchasing patterns have been employed over the years. However, this paper will primarily focus on methods such as RFM Analysis as well as Clustering Analysis, which will be elaborated on in subsequent sections.

RFM and Clustering analyses

RFM Analysis is a widely used technique for customer segmentation, behaviour analysis and determining lifetime value. According to Maraghi *et al.* (2020), RFM Analysis is an established and effective method for deriving insights from customer relationship management (CRM) data. Initially proposed by Hughes (1994), the RFM model is considered straightforward and efficient. Clustering Analysis, which includes model-based, grid-based, density-based, partition-based and hierarchical methods, is frequently used alongside RFM Analysis for segmenting customers (Christy *et al.*, 2021).

Xie *et al.* (2019) also noted that Clustering Analysis is crucial for feature identification and customer segmentation in data mining. Abdulhafedh (2021) found that K-Means Clustering was particularly appropriate for customer segmentation in their research and noted the effectiveness of Principal Component Analysis in reducing dataset dimensions.

Besides that, Idowu & Kattukottai (2019) compared three clustering algorithms: Hierarchical, K-Means and Fuzzy C-Means in their study to understand customer behaviour and segment them according to purchasing patterns using RFM Analysis. Their results indicated that while K-Means and Fuzzy C-Means identified five major clusters, the Hierarchical approach produced two clusters and demonstrated superiority with a Dunn Index score of 1.58. This

https://doi.org/10.18080/jtde.v12n3.978

model was found to be better suited for generating clusters with high inter-cluster distances and low intra-cluster distances.

Moreover, Anitha & Patil (2022) employed the K-Means algorithm in combination with RFM Analysis to analyse customer segmentation, behaviour and retention by using Silhouette Analysis to determine the optimal number of clusters. They concluded that obtaining a high Silhouette Score can offer businesses a robust framework for enhancing customer retention by examining sales' RFM values.

In a previous study conducted by Serwah *et al.* (2023), the focus was on enhancing customer segmentation through the use of RFM Analysis combined with K-Means and weighted K-Means Clustering. These methods categorise customers into groups such as "Need Attention", "Loyal Customers" and "At Risk", aiming to refine marketing strategies. The results demonstrated that weighted K-Means, which prioritises certain RFM factors, achieved better segmentation with a Silhouette Score of 0.40, compared to 0.30 for traditional K-Means. Despite its effectiveness, the study notes limitations related to data outliers and its focus on a specific demographic.

Differentiation and advances

This study differentiates itself by integrating RFM Analysis with Hierarchical Clustering, using innovative data pre-processing steps like log transformation and the Power Transformer with the Yeo-Johnson parameter to enhance segment accuracy and interpretability. Unlike previous research that focuses on a single clustering method, this work offers a comparative analysis using real-world e-commerce data and unique evaluation metrics, such as the Silhouette Score and Calinski–Harabasz Index. This approach provides valuable insights for developing targeted marketing strategies based on robust customer segmentation, setting this research apart within the field of e-commerce analytics.

Methodology

This section outlines the proposed methodology, detailing the dataset used and the processes underpinning the employed methods.

https://doi.org/10.18080/jtde.v12n3.978

Proposed methodology

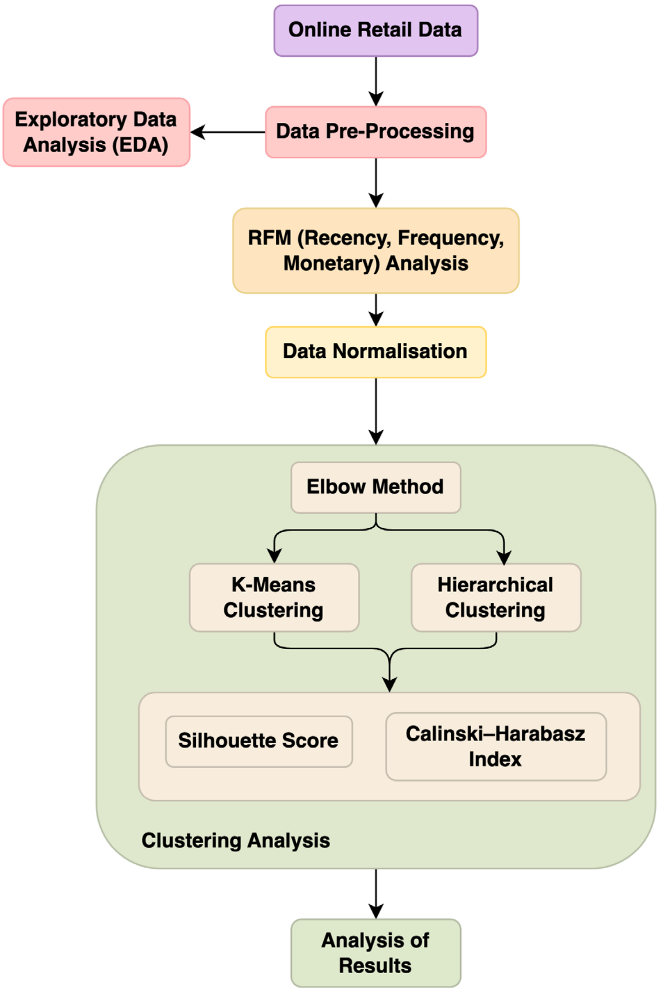


Figure 1. Conceptual Model of the Study

Dataset

This research utilises a dataset from the UCI Machine Learning Repository, a highly regarded source for machine learning datasets, consisting of 541,909 transactions. The dataset comprises transactional data from a UK-based online retail business, covering all transactions from 1 December 2010 to 9 December 2011 (Chen, 2015). The dataset was originally collected to study e-commerce transaction patterns and customer behaviour.

Additionally, the dataset's attributes such as InvoiceNo, StockCode, Description, Quantity, InvoiceDate, UnitPrice, CustomerID and Country are crucial for performing RFM Analysis and clustering. These attributes provide a comprehensive view of customer transactions, reflecting real-world e-commerce activities and making the dataset representative of typical e-commerce operations.

Furthermore, primary data collection was not conducted for this study because the existing, well-documented dataset aligns with our research objectives. This choice allowed us to focus on analysis rather than the logistics of primary data collection. The relevance of this dataset to real-world scenarios is crucial for studying customer segmentation, which is essential for developing targeted marketing strategies and enhancing business performance (Chen *et al.*, 2012). Table 1 outlines the specific attributes contained in the dataset.

Table 1. Attributes in the Online Retail Dataset

No	Attribute Name	Attribute Description	Data Type
1	InvoiceNo	A unique six-digit Invoice Number is assigned to each transaction made. Invoice Numbers that start with the letter "C" indicate that the transaction has been cancelled.	Nominal
2	StockCode	A five-digit unique number which is assigned to each discrete product.	Nominal
3	Description	The name of the product.	Nominal
4	Quantity	The number of each product in a single transaction.	Numeric
5	InvoiceDate	The time and date in which each transaction was made.	Numeric
6	UnitPrice	The unit price of each of the products in pounds (£).	Numeric
7	CustomerID	A five-digit unique number which is assigned to each customer.	Nominal
8	Country	The country of residence for each customer.	Nominal

Data pre-processing

In order to ensure data quality and accuracy, a thorough data pre-processing step was conducted before any analysis. Figure 1 illustrates the overall process from preprocessing to exploratory data analysis (EDA). Figure 2 provides a detailed breakdown of these preprocessing steps, which include handling duplicate values, addressing missing data, and transforming columns into the appropriate format for further analysis. Hence, to eliminate potential bias or distortion in the dataset, any duplicate entries were removed. Additionally, to address missing data, records without a CustomerID were excluded from the analysis, as this attribute was crucial for our study.

Moreover, an Exploratory Data Analysis (EDA) was conducted to investigate various aspects, including order cancellations, correlations between different attributes, discount offerings by the online retail company, and countries with the highest spending power. The primary objective of this analysis was to identify patterns or trends related to customer purchasing behaviour, the impact of discounts, and any other relevant insights. By examining the relevant variables, we aimed to gain a deeper understanding of the effectiveness and implications of the company's marketing strategies. These data pre-processing and EDA steps played a crucial

https://doi.org/10.18080/jtde.v12n3.978

role in ensuring the integrity of the data and facilitating meaningful and reliable analysis of the online retail dataset.

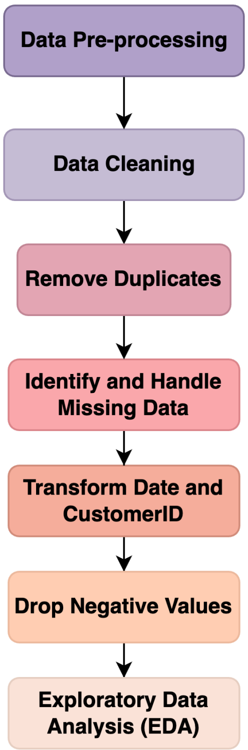


Figure 2. Data Pre-processing Steps

- Data Pre-processing: A critical preliminary step that encompasses all preparatory actions before any advanced analysis. This includes cleaning the data, transforming variables, and ensuring that the dataset is ready for further processing.
- Data cleaning: Removed duplicates to eliminate bias and distortion in the dataset. Duplicates can artificially inflate the frequency of transactions for certain customers, leading to incorrect segmentation.
- Handling missing data: Excluded records without customer ID as this attribute is crucial for identifying unique customers. Including records without this attribute would result in incomplete and misleading analysis.
- Transforming date and customer ID: Reformatted date and customer ID to ensure consistency and suitability for RFM and Clustering analyses. Correct formatting ensures accurate calculation of recency and proper customer identification.
- Dropping negative values: Removed any negative values to maintain data integrity. Negative values in transaction data can be due to returns or errors and can distort the analysis.

- Conducting EDA: Undertaken to uncover patterns and insights within the dataset. EDA helps in understanding the data distribution, identifying outliers and forming hypotheses for further analysis.

Overview of RFM Analysis

Firstly, for RFM Analysis, it is necessary to assign an RFM value to each customer in the dataset. It is important to emphasise that RFM Analysis can only be applied to customers who have a valid customer ID, which is why records without a customer ID were removed during the data pre-processing stage. Besides that, recency is determined by calculating the time difference between the current date and the most recent invoice date while frequency is calculated by counting the total number of transactions made by each customer. The monetary value is then computed by summing the total amount spent by a customer across all orders.

These calculations allow us to capture important aspects of customer behaviour and enable effective analysis and segmentation of customers based on their recency, purchasing frequency and monetary contributions. After calculating the RFM values, we assign RFM scores by segmenting the dataset into five equal parts, or quintiles, utilising quantile-based discretisation. This method is selected to generate scores from 1 to 5, offering a nuanced exploration of customer segments compared to a quartile-based scoring system, which limits scores to a range from 1 to 4 (Fahrudin & Rindiyani, 2024).

In this study, a recency score of 5 denotes customers who have engaged in recent purchases, whereas a score of 1 signifies those who have remained inactive over an extended period. Furthermore, with respect to frequency, customers achieving a score of 5 have engaged in more frequent transactions compared to their counterparts with a score of 1. Similarly, in terms of monetary value, a score of 5 is attributed to customers who have expended significantly more, contrasting with those assigned a score of 1. These scoring mechanisms allow for the differentiation of customer behaviours on the dimensions of RFM value, offering critical insights essential for the effective segmentation and formulation of targeted marketing strategies.

After merging the RFM scores for each customer based on their customer ID, customer segments are created by combining these scores into unified profiles for each customer group. According to Nair (2023), an author at Putler which is an online vendor specialising in analytical services, 11 distinct segments have been identified, each corresponding to unique combinations of the RFM metrics, providing insights into customer behaviour patterns. These segments are detailed in Table 2, illustrating the varied customer behaviours captured through the RFM scoring system.

https://doi.org/10.18080/jtde.v12n3.978

Table 2. RFM Scores and Corresponding Customer Segments

Segments	Recency	Frequency	Monetary
Champions	4–5	4–5	4–5
Loyal Customers	2–5	3–5	3–5
Potential Loyalists	3–5	1–3	1–3
Recent Customers	4–5	0–1	0–1
Promising	3–4	0–1	0–1
Customers Needing Attention	2–3	2–3	2–3
About To Sleep	2–3	0–2	0–2
At Risk	0–2	2–5	2–5
Can't Lose Them	0–1	4–5	4–5
Hibernating	1–2	1–2	1–2
Lost	0–2	0–2	0–2

Furthermore, given the successful utilisation of segment labels from Kabaskal (2020), Nair (2023) and Sutresno *et al.* (2018) in their case studies to define RFM segments, it is evident that adopting a similar approach in our research is feasible. This precedent demonstrates the applicability of Putler's segment labels in effectively categorising customer behaviours based on RFM criteria.

Clustering Analysis

In this section, we will explore clustering techniques, specifically Hierarchical and K-Means Clustering, focusing on the values derived from the RFM Analysis. However, to ensure fairness and eliminate biases in the data, we will begin by normalising the data using two techniques, which are the log transformation and Power Transformer with the Yeo-Johnson parameter.

By applying data normalisation techniques such as the log transformation and Power Transformer with the Yeo-Johnson parameter, it helps to address any skewness or non-normality in the data and ensures that each variable contributes proportionately to the analysis, which prevents any single variable from overpowering the others (Yeo, 2000). Thus, these data normalisation techniques can eliminate biases in the variables, enabling a more accurate comparison. However, the Power Transformer with the Yeo-Johnson parameter is more flexible, handling both positively and negatively skewed data more effectively. This choice enhances the suitability of the data for subsequent Clustering Analysis, ensuring accurate comparison and clustering results. Further details about this normalisation process are discussed in the Results and Discussion section, particularly within the data normalisation part, where a comprehensive understanding is provided.

Besides that, we will also be determining the optimal number of clusters which is represented by the k-value, for our clustering techniques. The Elbow Method will be employed as the primary approach for determining the optimal number of clusters. Additionally, evaluation metrics such as the Silhouette Score and Calinski–Harabasz Index will be used to assess the

https://doi.org/10.18080/jtde.v12n3.978

quality and effectiveness of the clustering, which are important in identifying the most suitable clusters for the data, ensuring that the results are accurate and meaningful. K-Means Clustering was chosen for its efficiency and scalability, making it suitable for large datasets and well-separated clusters. On the other hand, Hierarchical Clustering was selected for its flexibility in managing clusters of varying shapes, sizes and densities, and for its ability to reveal the data's underlying structure through dendrograms. Although Hierarchical Clustering does not initially require specifying the number of clusters, the optimal number of clusters (k) can be determined by analysing the dendrogram and choosing a level at which to cut the tree. This method is useful for exploratory analysis and understanding the hierarchical relationships within the data.

Table 3. Comparison of Hierarchical and K-Means Clustering Methods

Method	Parameters	Advantages	Uses
Hierarchical Clustering	Agglomerative, linkage criteria (e.g., ward, complete, average), number of clusters (k)	Flexible in managing clusters of varying shapes, sizes, and densities; provides a dendrogram for exploratory analysis	Suitable for understanding the hierarchical structure of smaller to moderately large datasets
K-Means Clustering	Number of clusters (k), initialisation method	Efficient and scalable for large datasets, straightforward implementation	Suitable for larger datasets, well-separated clusters

Results and Discussion

We will provide a comprehensive explanation of the analysis conducted and the resulting findings in detail. The results obtained from the analysis will be thoroughly discussed, providing a clear understanding of the outcomes and their implications. Through this discussion of the outcomes, we aim to provide a comprehensive overview of the analysis and its relevance to the research objectives.

Interpretation of RFM Segments

The findings revealed that 3,700 out of 3,921 customers were assigned to a segment. Among these, segments from eight of the eleven categories proposed by Nair (2023) and Sutresno *et al.* (2018) were identified as shown in Figure 3. However, the analysis also highlights three anticipated segments: Recent Customers, Promising, and Lost, that were not identified in the results. This discrepancy may be attributed to the restricted range of purchase dates within our dataset.

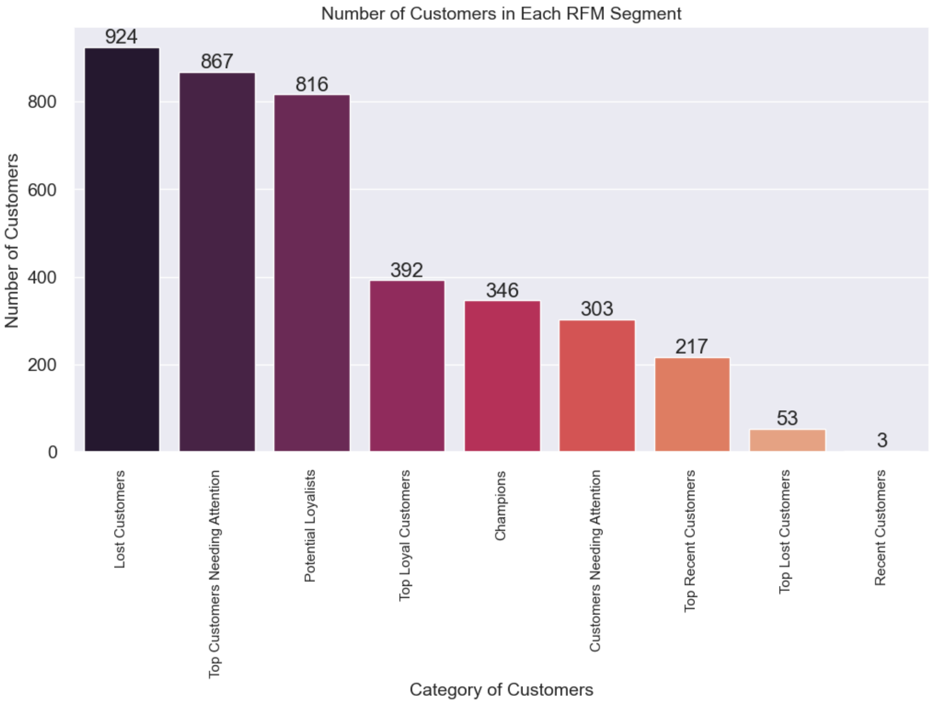


Figure 3. Bar Chart of Customer Distribution in RFM Segments

Moreover, as emphasised by Kabaskal (2020), data diverging from established clusters is commonly identified as outliers, which do not conform to any specific customer segment. Consequently, the remaining customers, labelled as "Others" may be classified as exceptions. It is essential to note that these customers belong to an RFM segment different from the ones discussed by Nair (2023) and Sutresno *et al.* (2018). Thus, to ensure the robustness and interpretability of our clustering results, customers falling under the "Others" label will be excluded from our analysis.

By omitting the "Others" label, we aim to mitigate the potential distortion of cluster centroids and the merging of clusters, which may arise from including outliers in the clustering process. This approach aligns with best practices in Clustering Analysis, where outliers are commonly treated separately to prevent biased or less interpretable results. Hence, we have a total of 8 segments, as illustrated in Table 4.

Table 4. Distribution of Customers Across RFM Segments

Segments	Recency
Loyal Customers	1237
Potential Loyalists	852
Champions	577
At Risk	412
Hibernating	331
About To Sleep	158
Customers Needing Attention	125
Can't Lose Them	8

https://doi.org/10.18080/jtde.v12n3.978

Data normalisation

Before applying the clustering techniques, we first explored different data normalisation techniques to address the skewed distribution of the data. We specifically considered log transformation and Power Transformer with the Yeo-Johnson parameter. The log transformation is commonly used to reduce skewness in positively skewed data. However, during our analysis, we noticed that some variables exhibited significant skewness, and the log transformation did not adequately normalise the data. Therefore, in order to further improve the normalisation process, we applied the Power Transformer technique with the Yeo-Johnson parameter. This method has the advantage of being able to handle both positively and negatively skewed data (Yeo, 2000), making it more flexible for our dataset.

We have observed that the Power Transformer with the Yeo-Johnson parameter yielded superior results in terms of normalising the data compared to the log transformation and has successfully reduced skewness and achieved a more symmetrical distribution, which enhances the suitability of the data for subsequent Clustering Analysis. This finding is illustrated in Figures 4 and 5, which show the improved normalisation that is achieved through the Power Transformer method with the Yeo-Johnson parameter, which outperforms the results obtained from the log transformation. Following that, we will then apply our transformed data using Hierarchical and K-Means Clustering.

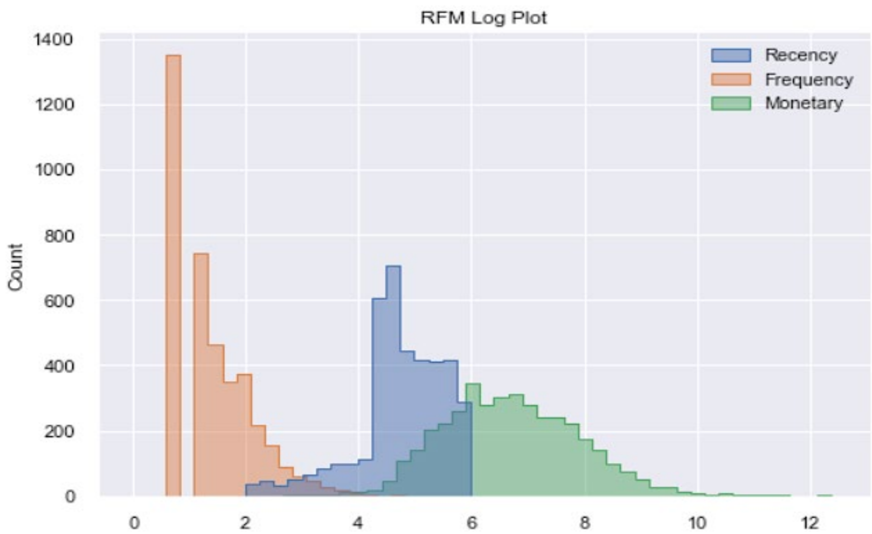


Figure 4. RFM Value Distribution using Log Transformation

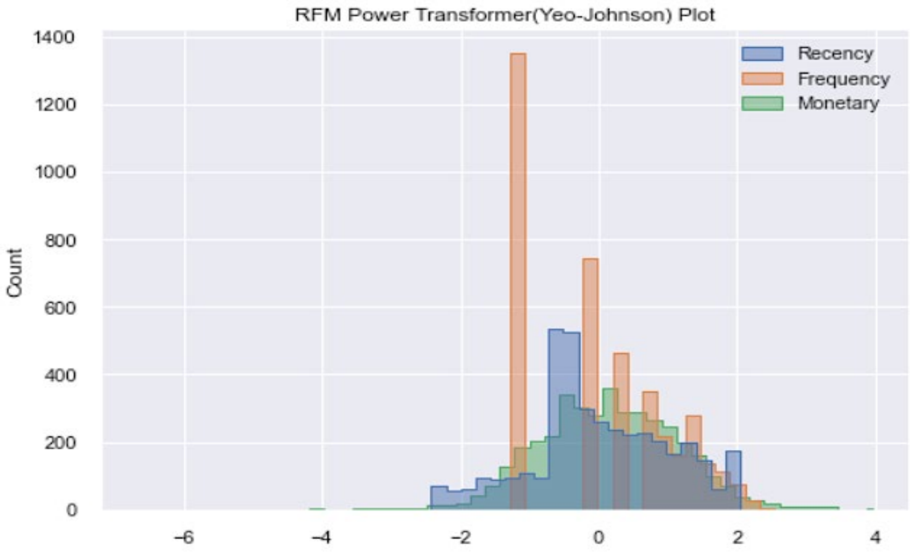


Figure 5. RFM Value Distribution using Power Transformer (Yeo-Johnson)

RFM with Hierarchical (Agglomerative) Clustering

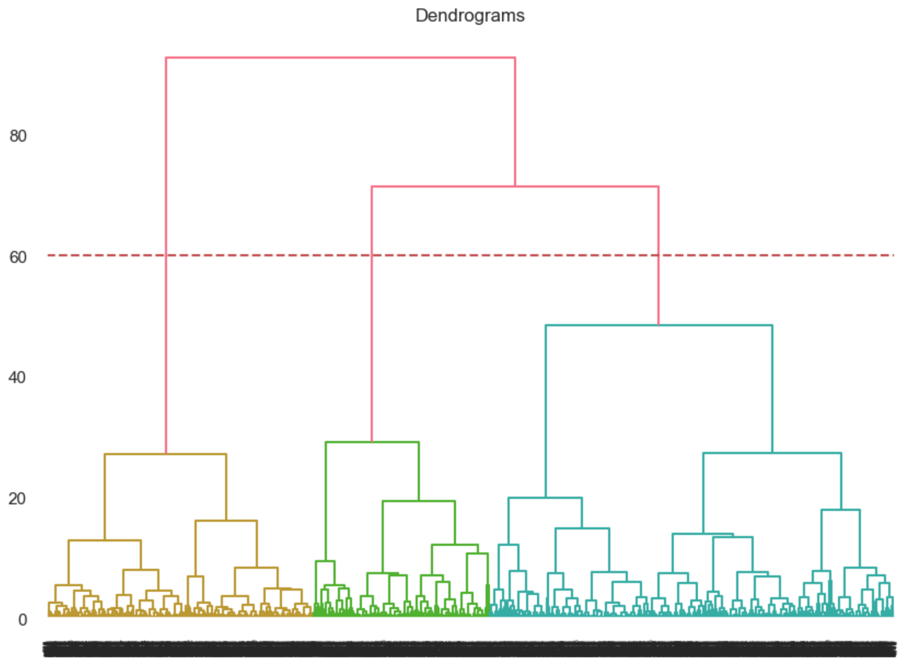


Figure 6. Dendrogram for RFM-Hierarchical (Agglomerative) Clustering

To determine the optimal number of clusters for Hierarchical (Agglomerative) Clustering, we applied the Elbow Method alongside the analysis of dendrograms. Figure 6 displays the dendrogram, where the horizontal line intersects three vertical lines, indicating a natural division into three clusters. Complementary to this, Figure 7 presents the Elbow Method plot, which shows a distinct "elbow" at k = 3. These figures collectively lead us to conclude that three is the optimal number of clusters for our analysis. The dendrogram was generated using the Euclidean distance measure and the Ward linkage method, which minimises the variance within each cluster. A cut-off distance of 60 was chosen based on the visual inspection of the dendrogram. This distance provides a clear separation of clusters, ensuring distinct and meaningful groupings.

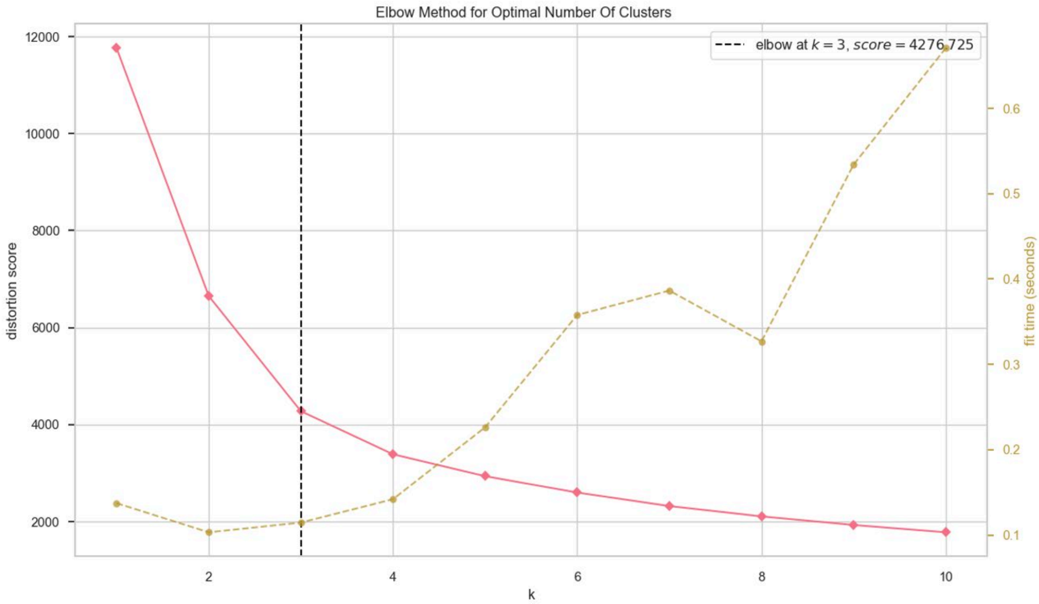


Figure 7. Elbow Method for Determining Optimal Clusters in RFM-Hierarchical (Agglomerative) Clustering

The choice of the Ward method enhances the compactness and homogeneity of clusters, making them more useful for customer segmentation and targeted marketing strategies. The Ward linkage method is particularly effective for creating well-defined clusters, as it considers the variance within clusters, leading to more accurate and interpretable results (Everitt & Hothorn, 2011; Kassambara, 2017).

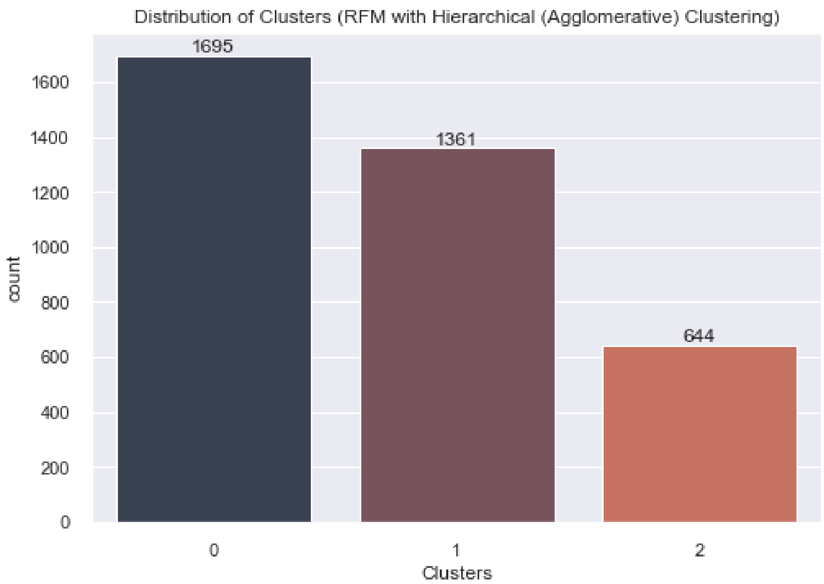


Figure 8. Cluster Distribution using RFM-Hierarchical (Agglomerative) Clustering

It is noteworthy that the x-axis labels on the dendrogram are not clearly visible because of the large dataset size, comprising a substantial number of data points. However, this does not detract from the dendrogram's effectiveness in illustrating the clustering structure and the chosen cut-off point, which robustly represents the hierarchical relationships within the data.

After initialising the RFM-Hierarchical (Agglomerative) Clustering model, we will proceed to fit the model to the data and make predictions regarding the clusters. Figure 8 illustrates the result of applying Hierarchical (Agglomerative) Clustering, where three distinct clusters have been derived.

Following this, we will conduct a detailed examination of the customer segments within each identified cluster. The dataframe shown in Figures 9 to 11 display distinct customer segments within each of the three clusters.

CustomerID	RFM	RFM_Score	custCategory	ClusterIDAgglo
17790	355	13	Loyal Customers	0
17667	245	11	Loyal Customers	0
16136	443	11	Loyal Customers	0
16464	444	12	Champions	0
14609	443	11	Loyal Customers	0

Figure 9. Sample Dataframe for Customers in Cluster 0 (Hierarchical Clustering)

CustomerID	RFM	RFM_Score	custCategory	ClusterIDAgglo
17698	121	4	Hibernating	1
13302	211	4	About To Sleep	1
17715	222	6	Customers Needing Attention	1
16473	242	8	At Risk	1
14729	112	4	Hibernating	1

Figure 10. Sample Dataframe for Customers in Cluster 1 (Hierarchical Clustering)

CustomerID	RFM	RFM_Score	custCategory	ClusterIDAgglo
14968	511	7	Potential Loyalist	2
15937	511	7	Potential Loyalist	2
15592	512	8	Potential Loyalist	2
17977	522	9	Potential Loyalist	2
13349	512	8	Potential Loyalist	2

Figure 11. Sample Dataframe for Customers in Cluster 2 (Hierarchical Clustering)

Furthermore, we have also visualised the RFM variables for each customer cluster obtained through Hierarchical Clustering using a boxplot shown in Figure 12. By combining the insights from the dataframe with the RFM boxplots for each cluster, we can then categorise the customers into clusters. These categorisations are comprehensively outlined in Table 5.

Table 5. Predicted Clusters (Hierarchical Clustering) and their Customer Description

Cluster	Customer Description	Number of Customers
0	Champions/Loyal Customers	1,695
1	About To Sleep/Hibernating/At Risk/Customers Needing Attention/Can't Lose Them	1,361
2	Potential Loyalists	644

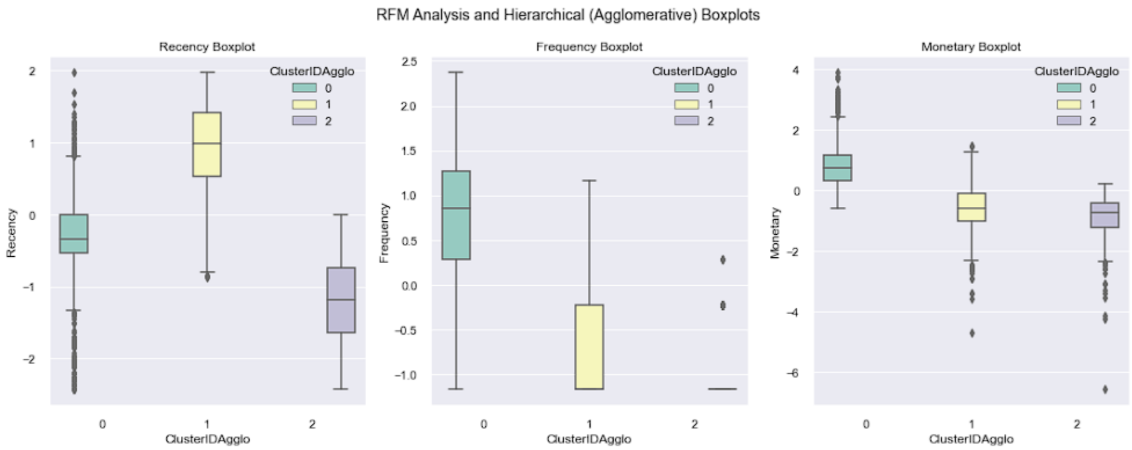


Figure 12. Boxplots for RFM with Hierarchical (Agglomerative) Clustering

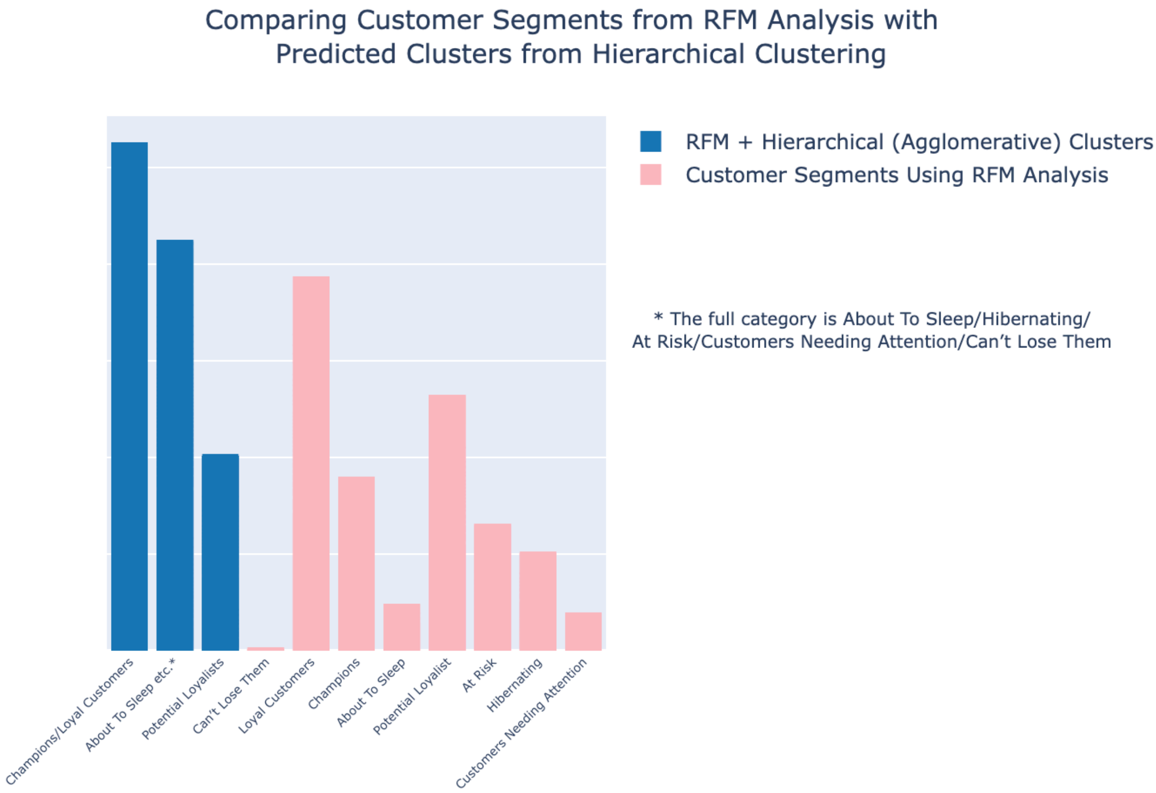


Figure 13. Comparison of Cluster Outcomes: RFM with Hierarchical Clustering vs Standalone RFM Analysis

https://doi.org/10.18080/jtde.v12n3.978

Additionally, Figure 13 presents a bar plot that compares the customer segments derived from integrating RFM Analysis with Hierarchical (Agglomerative) Clustering against those obtained from independently conducted RFM Analysis. This comparative assessment aims to evaluate the efficacy of combining clustering methods with RFM Analysis and to understand the implications of each approach on the segmentation results.

RFM with K-Means Clustering

Similar to our approach with Hierarchical Clustering, we will first identify the optimal number of clusters using the Elbow Method. As demonstrated in Figure 14, an "elbow" appears in the plot when the model is configured with three clusters, indicating that this is the optimal number of clusters for our dataset.

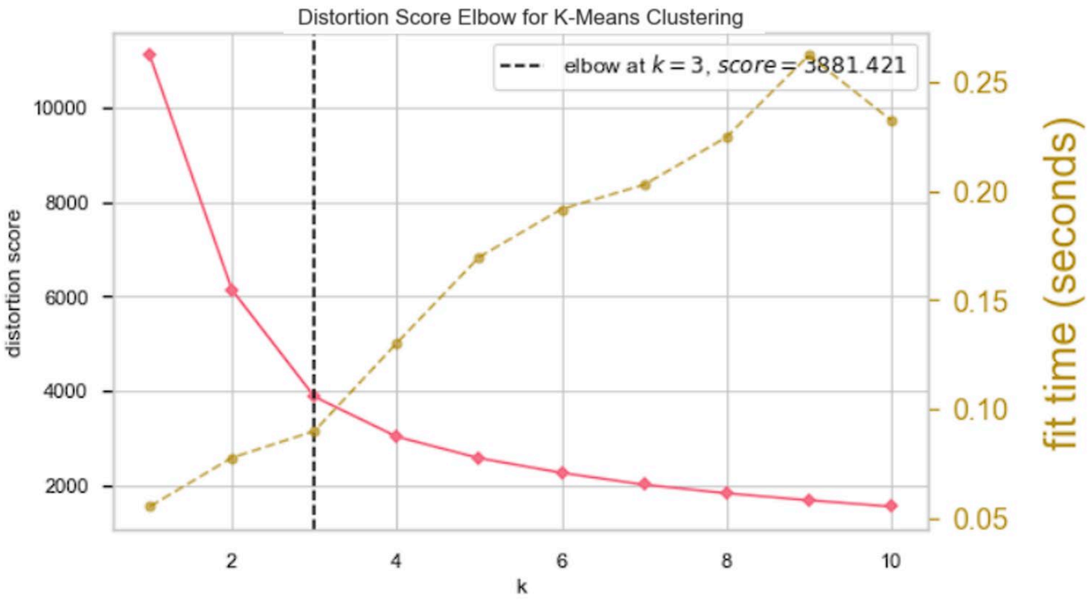


Figure 14. Elbow Method for Determining Optimal Clusters in K-Means Clustering

Besides that, we have also employed a Silhouette Visualiser to visualise and evaluate different values for the k-value by comparing various visualisers for k-values such as 2, 3 and 4 to strengthen our decision regarding the optimal number of clusters shown in Figures 15 to 17.

From Figures 15 to 17, it is evident that k = 3 outperforms the configurations where k = 2 and k = 4, as it exhibits a higher silhouette coefficient. Thus, these figures clearly suggest that the optimal number of clusters is three, aligning with the findings obtained from Hierarchical Clustering. Figure 18 illustrates the distribution of the three clusters identified through K-Means Clustering.

https://doi.org/10.18080/jtde.v12n3.978

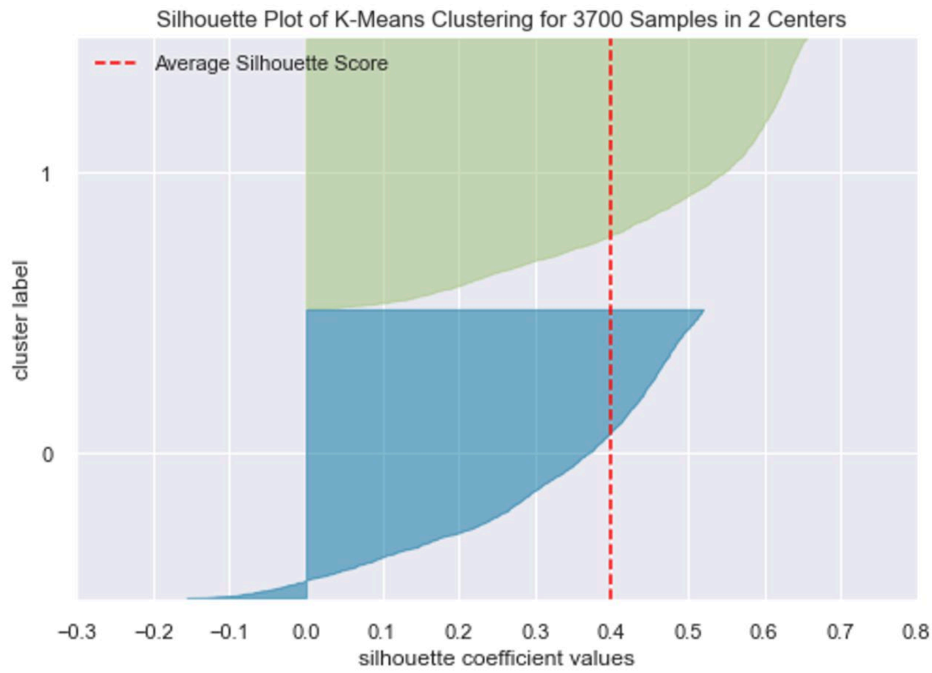


Figure 15. Silhouette Visualiser for k = 2

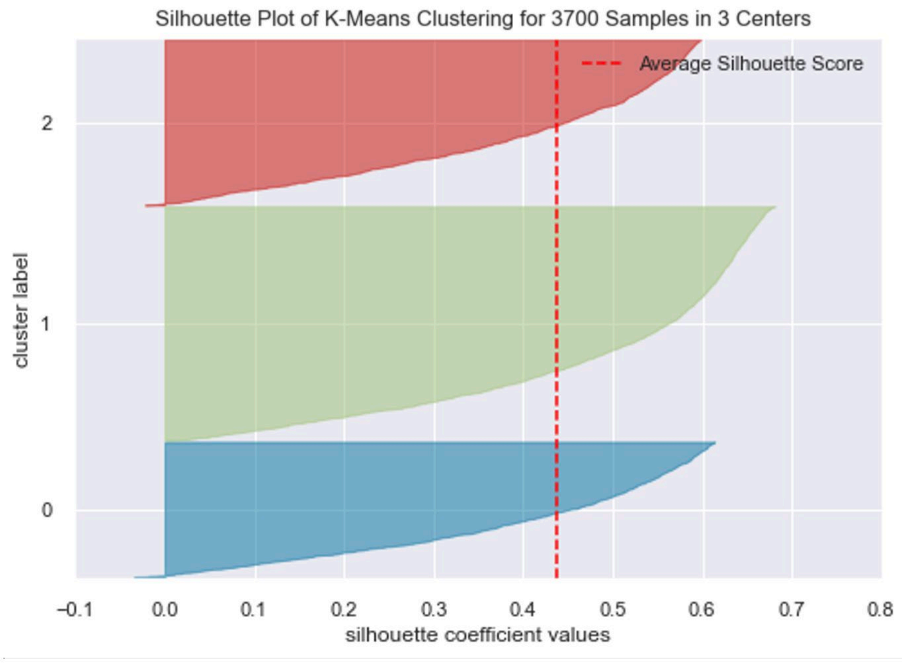


Figure 16. Silhouette Visualiser for k = 3

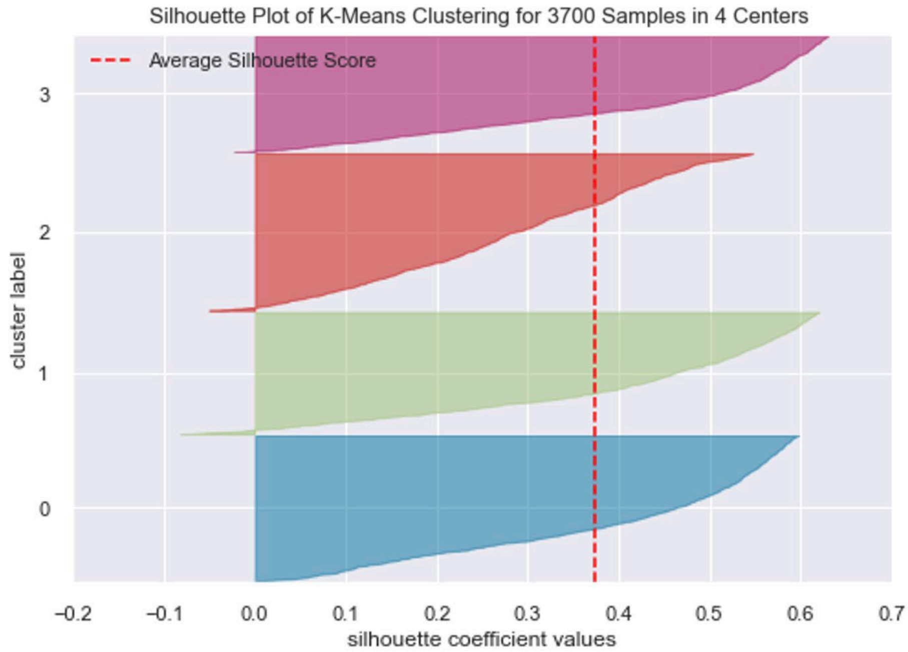


Figure 17. Silhouette Visualiser for k = 4

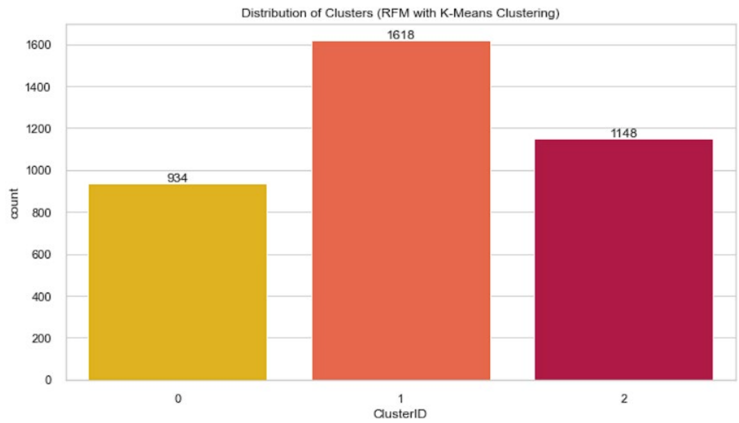


Figure 18. Cluster Distribution using RFM with K-Means Clustering

In our analysis, the RFM variables for each cluster, derived from K-Means Clustering, are depicted through the dataframe shown in [Figures 19]to [21], showcasing the distinct customer segments within each cluster.

CustomerID	RFM	RFM_Score	custCategory	ClusterID
18005	531	9	Potential Loyalist	0
15417	413	8	Potential Loyalist	0
16183	522	9	Potential Loyalist	0
17614	522	9	Potential Loyalist	0
18019	521	8	Potential Loyalist	0

Figure 19. Sample Dataframe for Customers in Cluster 0 (K-Means Clustering)

CustomerID	RFM	RFM_Score	custCategory	ClusterID
13113	355	13	Loyal Customers	1
17750	454	13	Champions	1
13107	444	12	Champions	1
14529	433	10	Loyal Customers	1
13471	545	14	Champions	1

Figure 20. Sample Dataframe for Customers in Cluster 1 (K-Means Clustering)

CustomerID	RFM	RFM_Score	custCategory	ClusterID
15709	111	3	Hibernating	2
13934	114	6	Can't Lose Them	2
13453	222	6	Customers Needing Attention	2
16990	221	5	About To Sleep	2
15589	133	7	At Risk	2

Figure 21. Sample Dataframe for Customers in Cluster 2 (K-Means Clustering)

Similar to the approach used with Hierarchical Clustering, we have visualised the RFM variables for each cluster obtained through K-Means Clustering using a boxplot. This visualisation enables us to categorise the three clusters as follows. Additionally, the distinctions between the clusters are further highlighted through a detailed boxplot, as illustrated in Figure 22. This visual representation allows for a clearer understanding of the variability and defining characteristics of each cluster.

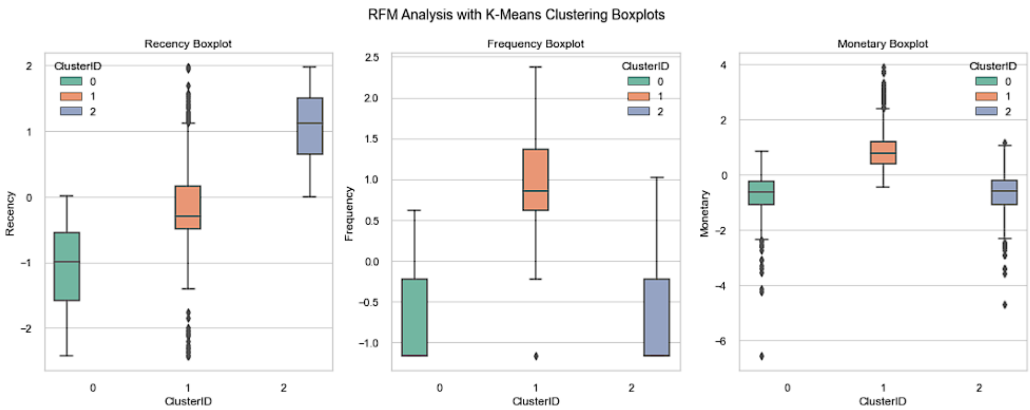


Figure 22. Boxplots for RFM with K-Means Clustering

The analysis and categorisation of these clusters, derived from the integration of both the dataframe and the boxplot is comprehensively outlined in Table 6.

Table 6. Predicted Clusters (K-Means Clustering) and Their Customer Description

Cluster	Customer Description	Number of Customers
0	Potential Loyalists	934
1	Champions/Loyal Customers	1,618
2	About To Sleep/Hibernating/At Risk/Customers Needing Attention/ Can't Lose Them	1,148

Subsequently, we performed a comparative analysis of the integrated approach of RFM with K-Means Clustering against the sole use of RFM Analysis. Figure 23 illustrates the customer clusters' distribution, offering a visual comparison of the results from both methodologies.

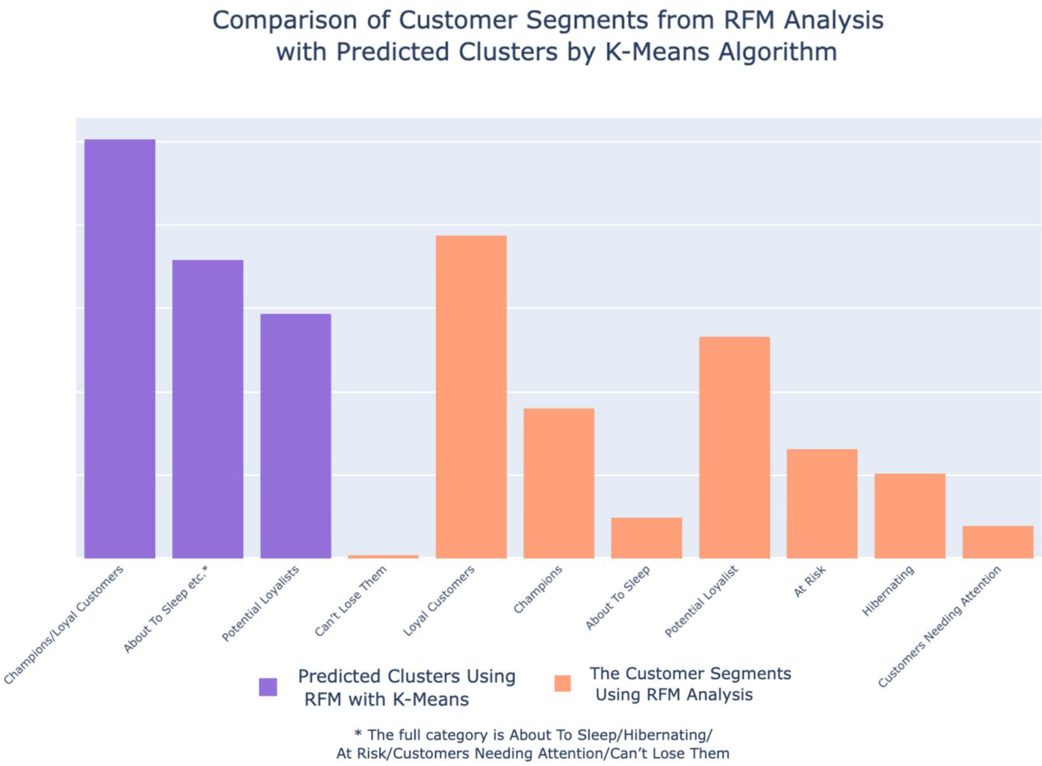


Figure 23. Comparison of Cluster Outcomes: RFM with K-Means Clustering vs Standalone RFM Analysis

The findings revealed that similar clusters were generated through RFM Analysis using both Hierarchical (Agglomerative) Clustering and K-Means Clustering techniques. Despite the interchange of cluster labels between the approaches, it is evident that both RFM with K-Means Clustering and RFM with Hierarchical Clustering have successfully generated three clusters of similar sizes. Furthermore, the underlying results closely align, demonstrating the robustness of the clustering regardless of the method used. In both methods, we identify equivalent clusters: "Champions/Loyal Customers" which represents our most valuable customers, "About to Sleep/Hibernating/At Risk/Customers Needing Attention/Can't Lose Them" denotes those with declining engagement, and "Potential Loyalists" indicates customers showing promise for increased loyalty. This alignment underscores the robustness of our segmentation approach and provides clear directions for focusing marketing efforts and strategies to enhance customer retention.

These findings suggest that both clustering techniques effectively and consistently identify distinct customer segments without significant impact from the choice of clustering method. However, the integrated approach, which results in a smaller number of customer segments compared to using RFM Analysis alone, provides advantages by identifying broader clusters with similar behaviours. This facilitates the development of more targeted and focused marketing strategies. In contrast, relying solely on RFM Analysis yields numerous distinct and separate clusters, complicating the segmentation and making it challenging to interpret. This complexity can obscure meaningful patterns and the unique characteristics of each segment.

However, practical considerations also play a role in choosing the appropriate clustering technique. Hierarchical Clustering, while providing detailed insights and higher interpretability through dendrograms, is computationally more demanding and better suited for smaller to moderately large datasets or detailed customer segmentation tasks. In contrast, K-Means Clustering stands out in scenarios requiring higher computational efficiency and scalability, such as in large-scale e-commerce applications, despite possibly yielding slightly lower Silhouette Scores.

By adopting the integrated approach, which groups customers into only three clusters based on similar behaviour, marketing strategies can be more easily customised for each cluster. This approach allows for a more streamlined and effective method of customer segmentation and targeted marketing efforts. Additionally, the emphasis placed on customer segments and marketing approaches can sometimes change, shaped by the company's resources, strategic goals and limitations.

Evaluation metrics

For the evaluation of the clusters, we will use the Silhouette Score and Calinski–Harabasz Index to assess the quality and effectiveness of the clustering as shown in Table 7. These metrics are crucial in identifying the most suitable clusters for the data, ensuring that the results are both accurate and meaningful.

- Silhouette Score: This metric measures how similar an object is to its own cluster compared to other clusters. It ranges from -1 to 1, where a higher score indicates better-defined clusters. This score considers both cohesion (how closely related objects in a cluster are) and separation (how distinct a cluster is from other clusters), making it ideal for assessing clustering quality (Rousseeuw, 1987).
- Calinski–Harabasz Index: Also known as the Variance Ratio Criterion, this index evaluates the ratio of the sum of between-cluster dispersion to within-cluster dispersion. A higher score indicates better-defined clusters. This metric measures the

compactness and separation of the clusters, ensuring they are distinct and well separated (Caliñski & Harabasz, 1974).

Table 7. Comparison of Clustering Performance Metrics: Silhouette Score and Calinski–Harabasz Index for RFM with K-Means and Hierarchical Clustering

	Silhouette Score	Calinski–Harabasz Index
RFM with K-Means Clustering	0.44	3437.8
RFM with Hierarchical (Agglomerative) Clustering	0.47	3787.1

The observed Silhouette Scores, although modest, are consistent with findings from previous research. For instance, a study by Serwah *et al.* (2023) reported a Silhouette Score of 0.40 for weighted K-Means and 0.30 for traditional K-Means, highlighting similar challenges in clustering performance. Their study also noted issues such as data outliers and a specific demographic focus, which impacted their results. Thus, our scores, while not exceptionally high, are within a reasonable range given the dataset's complexity and inherent limitations.

Besides that, the evaluation metrics presented in Table 7 indicate that RFM combined with Hierarchical Clustering outperforms RFM with K-Means Clustering in terms of the Silhouette Score and Calinski–Harabasz Index. This suggests that Hierarchical Clustering offers superior cluster quality, cohesion, and separation in creating three distinct clusters, making it a preferable option for devising effective marketing strategies. RFM with Hierarchical Clustering exhibits greater adaptability in managing clusters varying in shape, size and density. This adaptability ensures a more precise depiction of the dataset's inherent structure, which is particularly advantageous from a marketing standpoint. Here, understanding customer segments is crucial as they often display varied characteristics and behaviours. Additionally, higher evaluation scores imply that the clusters formed are well defined, with clear boundaries and minimal overlap.

Practical implications

The findings from our analysis have significant real-world applications for businesses seeking to enhance their marketing strategies through effective customer segmentation. By understanding the distinct customer segments identified through RFM Analysis and clustering techniques, businesses can tailor their marketing efforts to better meet the needs and preferences of each group. Table 8 outlines strategies and examples for each of the customer segments, demonstrating how businesses can implement these insights to drive customer engagement, satisfaction and loyalty, ultimately contributing to business growth and success.

https://doi.org/10.18080/jtde.v12n3.978

Table 8. Practical Implications of Customer Segmentation

Customer Segment	Recommended Strategy	Example
Champions/Loyal Customers	Offer exclusive rewards, loyalty programs, and personalised experiences to retain high-value customers.	Implement a loyalty program offering discounts, early access to new products and invitations to special events. This enhances satisfaction and repeat purchases.
Potential Loyalists	Engage with personalised marketing campaigns to encourage repeat purchases and foster brand loyalty.	Send targeted email campaigns highlighting products similar to previous purchases, offering special discounts or promotions to encourage further engagement.
About to Sleep/Hibernating/At Risk/Customers Needing Attention/Can't Lose Them	Re-engage through special offers, targeted promotions or personalised communication to prevent churn.	Utilise personalised email campaigns offering discounts on previously viewed or related products, along with personalised messages aimed at re-engaging customers.

Discussion

This study has successfully tackled the problem of identifying the most suitable clustering technique for recommending effective marketing strategies based on customer behaviour and purchasing patterns. The findings indicate that integrating RFM Analysis with Hierarchical (Agglomerative) Clustering achieves higher scores in evaluation metrics, which allows for a more accurate representation of the underlying structure of the dataset. This flexibility is particularly beneficial from a marketing perspective and is deemed essential for producing actionable recommendations.

Our study also demonstrated the significance of employing data normalisation methods for the RFM values, such as the Power Transformer with the Yeo-Johnson parameter. These methods ensure an accurate representation of customer behaviour and enhance the reliability and interpretability of the clustering results. Overall, this research contributes to the existing literature by addressing gaps related to the comparison and interpretability of RFM Analysis with clustering techniques, the importance of data normalisation and the practical implementation of discovered customer segments to enhance marketing strategies. The insights obtained from this study provide valuable guidance for businesses in developing customised marketing approaches, enhancing customer satisfaction and driving revenue growth.

Conclusion

In conclusion, this study has generated new knowledge on effective approaches for enhancing marketing strategies through customer segmentation based on behavioural analysis. By leveraging RFM Analysis with Hierarchical (Agglomerative) Clustering, businesses can

https://doi.org/10.18080/jtde.v12n3.978

identify distinct customer segments and tailor their marketing strategies to target these groups more effectively. This research adds significant value to the field of e-commerce analytics by demonstrating the practical application and benefits of integrating RFM Analysis with Hierarchical Clustering, providing actionable insights for personalised marketing strategies.

Key findings

- RFM with Hierarchical Clustering provides better segmentation accuracy and detail for smaller datasets, as evidenced by superior Silhouette Scores and Calinski–Harabasz Index values.
- The integrated approach of RFM Analysis with clustering techniques offers a streamlined method for customer segmentation, leading to more targeted and effective marketing strategies.
- Data normalisation, especially using the Power Transformer with the Yeo-Johnson parameter, significantly impacts clustering results, enhancing the accuracy and interpretability of customer segments.

Limitations

While this study offers valuable insights, it has several limitations. The findings are based on a specific dataset, which may not be generalisable across all industries or customer demographics. Furthermore, the scope was restricted to traditional RFM metrics, without considering more dynamic customer interaction variables. These constraints underscore the necessity for additional research to validate and extend the results.

Future research directions

Future studies should consider:

- Additional variables such as demographic information, social media activity and customer feedback to gain a more comprehensive view of customer behaviour and preferences
- Investigating advanced clustering methods such as density-based or spectral clustering for more nuanced insights
- Focusing on integrating machine learning with clustering to predict customer behaviour changes in real time, enhancing the adaptability of marketing strategies.

Despite its limitations, this study contributes valuable perspectives to the field of marketing by suggesting potential approaches for crafting targeted marketing strategies based on customer behaviour. By understanding and utilising these customer segments, businesses can enhance their marketing efforts, potentially leading to improved customer satisfaction and business growth.

Acknowledgements

I, Wong Chun Gee, as the main author, would like to extend my sincere appreciation to my supervisors, who are also my co-authors, for their invaluable support and guidance throughout this research. Their insights and discussions have greatly enriched my understanding of the topic.

References

Abdulhafedh, A. (2021). Incorporating K-means, hierarchical clustering and PCA in customer segmentation. *Journal of City and Development*, *3*(1), 12–30. ResearchGate. https://www.researchgate.net/publication/349094412_Incorporating_K-means_Hierarchical_Clustering_and_PCA_in_Customer_Segmentation

Anitha, P., & Patil, M. M. (2022). RFM model for customer purchase behavior using K-Means algorithm. *Journal of King Saud University – Computer and Information Sciences*, *34*(5), 1785–1792. https://doi.org/10.1016/j.jksuci.2019.12.011

Caliñski, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics*, *3*(1), 1–27. https://doi.org/10.1080/03610927408827101

Chaubey, G., Gavhane, P. R., Bisen, D., & Arjaria, S. K. (2022). Customer purchasing behavior prediction using machine learning classification techniques. *Journal of Ambient Intelligence and Humanized Computing*. https://doi.org/10.1007/s12652-022-03837-6

Chen, D. (2015). Online retail. *UC Irvine Machine Learning Repository*. https://doi.org/10.24432/C5BW33

Chen, D., Sain, S. L., & Guo, K. (2012). Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining. *Journal of Database Marketing and Customer Strategy Management*, *19*(3), 197–208. https://doi.org/10.1057/dbm.2012.17

Christy, A. J., Umamakeswari, A., Priyatharsini, L., & Neyaa, A. (2021). RFM ranking – An effective approach to customer segmentation. *Journal of King Saud University – Computer and Information Sciences*, *33*(10), 1251–1257. https://doi.org/10.1016/j.jksuci.2018.09.004

Everitt, B., & Hothorn, T. (2011). *An Introduction to Applied Multivariate Analysis with R*. Springer. https://www.springer.com/series/6991

Fahrudin, N. F., & Rindiyani, R. (2024). Comparison of k-medoids and k-means algorithms in segmenting customers based on RFM criteria. *E3S Web of Conferences*, *484*. https://doi.org/10.1051/e3sconf/202448402008

Garg, A., Popli, R., & Sarao, B. S. (2021). Growth of digitization and its impact on big data analytics. *IOP Conference Series: Materials Science and Engineering*, *1022*(1). https://doi.org/10.1088/1757-899X/1022/1/012083

Griva, A., Bardaki, C., Pramatari, K., & Papakiriakopoulos, D. (2018). Retail business analytics: Customer visit segmentation using market basket data. *Expert Systems with Applications*, *100*, 1–16. https://doi.org/10.1016/j.eswa.2018.01.029

Hughes, A. M. (1994). *Strategic Database Marketing*. Probus Publishing.

Idowu, S., & Kattukottai, S. (2019). Customer segmentation based on RFM model using k-means, hierarchical and fuzzy c-means clustering algorithms. ResearchGate. https://doi.org/10.13140/RG.2.2.15379.71201

Kabaskal, İ. (2020). Customer segmentation based on recency frequency monetary model: A case study in e-retailing. *Bilişim Teknolojileri Dergisi*, *13*(1), 47–56. https://doi.org/10.17671/gazibtd.570866

Kadir, M., & Achyar, A. (2019) *Customer segmentation on online retail using RFM analysis: Big data case of Bukku.id*. European Union Research Library. https://doi.org/10.4108/eai.1-4-2019.2287279

Kassambara, A. (2017). *Practical guide to cluster analysis in R: Unsupervised machine learning*. https://xsliulab.github.io/Workshop/2021/week10/r-cluster-book.pdf

Kaur, M., & Aggarwal, K. (2022). *Big data market trends and its impact on e-commerce industry*. ResearchGate. https://www.researchgate.net/publication/359024503_BIG_DATA_MARKET_TRENDS_AND_ITS_IMPACT_ON_E-COMMERCE_INDUSTRY

Kurniawan, F., Umayah, B., Hammad, J., Mardi, S., Nugroho, S., & Hariadi, M. (2018). Market basket analysis to identify customer behaviors by way of transaction data. *Knowledge Engineering*, *1*(1), 20–25. https://doi.org/10.17977/um018v1i12018p20-25

Li, X. (2021). Business analytics in E-commerce: a literature review. *Journal of Industrial Integration and Management*, *6*(1), 31–52. https://doi.org/10.1142/S2424862220500207

Mallam, P., Ashu, & Singh, B. (2021). Business intelligence techniques using data analytics: an overview. *Proceedings – 2021 International Conference on Computing Sciences, ICCS 2021*, 265–267. https://doi.org/10.1109/ICCS54944.2021.00059

Maraghi, M., Amin Adibi, M., & Mehdizadeh, E. (2020). Using RFM model and market basket analysis for segmenting customers and assigning marketing strategies to resulted segments. *Journal of Applied Intelligent Systems & Information Sciences*. https://doi.org/10.22034/jaisis.2020.102488

Mehmeti, G., & Luga, E. (2021). The influence of situational factors on consumer purchasing behavior – the case of Covid-19. *Albanian Journal of Agricultural Sciences, 20*(2).

Nair, A. (9 November 2023). *RFM analysis for successful customer segmentation*. Putler. https://www.putler.com/rfm-analysis/

Raguseo, E., & Vitari, C. (2018). Investments in big data analytics and firm performance: an empirical investigation of direct and mediating effects. *International Journal of Production Research*, *56*(15), 5206–5221. https://doi.org/10.1080/00207543.2018.1427900

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. In *Journal of Computational and Applied Mathematics*, *20*. https://doi.org/10.1016/0377-0427(87)90125-7.

Sarkar, M., Aisharyja, Puja, R., & Chowdhury, F. R. (2024). Optimizing marketing strategies with RFM method and k-means clustering-based AI customer segmentation analysis. *Journal of Business and Management Studies*. https://doi.org/10.32996/jbms.2024.6.2.5

Serwah, A., Khaw, K. W., Cheang, S. P. Y., & Alhamzah, A. (2023). Customer analytics for online retailers using weighted k-means and RFM analysis. *Data Analytics and Applied Mathematics*, 1–7. https://doi.org/10.15282/daam.v4i1.9171

Sutresno, S. A., Iriani, A., & Sediyono, E. (2018). Metode K-Means Clustering dengan Atribut RFM untuk Mempertahankan Pelanggan. *JuTISI*, *4*(3), 433–440. Retrieved from https://journal.maranatha.edu/index.php/jutisi/article/view/1479

Xie, H., Zhang, L., Lim, C. P., Yu, Y., Liu, C., Liu, H., & Walters, J. (2019). Improving K-means clustering with enhanced Firefly algorithms. *Applied Soft Computing Journal*, *84*. https://doi.org/10.1016/j.asoc.2019.105763

Yeo, I.-K. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika*, *87*(4), 954–959. https://doi.org/10.1093/biomet/87.4.954

https://doi.org/10.18080/jtde.v12n3.978

Deep Reinforcement Learning in Cryptocurrency Trading: A Profitable Approach

Xue Hao Tay
Faculty of Computing and Information Technology, Tunku Abdul Rahman University of Management and Technology, Malaysia

Siew Mooi Lim
Faculty of Computing and Information Technology, Tunku Abdul Rahman University of Management and Technology, Malaysia

Abstract: This study proposes an Automatic Cryptocurrency Trading System using Deep Reinforcement Learning (DRL). Six popular cryptocurrencies were used: Bitcoin, Ethereum, BinanceCoin, DogeCoin, Cardano, and WAVES. Development of the trading system started with building three timeseries models – Temporal Convolutional Neural Network (TCNN), Long Short-Term Memory Network (LSTM), and Gated Recurrent Unit Network (GRU) – to predict future prices. Then, cryptocurrency sentiment data was scraped using the Alternative.me API. Data on historical prices, predicted future prices, cryptocurrency sentiment index, technical indicators, and trading account information was fed as input states to three DRL Agents — Deep Q Network (DQN), Advantage Actor Critic (A2C), and Recurrent Proximal Policy Optimization (RPPO) — which were trained using a custom-developed trading environment. Each agent was given $1000 initial capital for all six cryptocurrencies to trade using three possible actions — Buy, Sell and Hold — and were back-tested on one year of unseen data. Our DQN model had the highest overall return on investment (ROI) of $740, an average 12.3% ROI across all six cryptocurrencies, with an ROI of 63.98% achieved for BinanceCoin. However, A2C and RPPO both had negative ROI.

Keywords: Investment, Inflation, Cryptocurrency, Timeseries, Deep Reinforcement Learning

Introduction

The Department of Statistics Malaysia has reported that Malaysia's inflation rate has risen to 1.8% year-on-year as of February 2024 which is an increase compared to January 2024 and December 2023, which were both at 1.5% (DOSM, 2024a; 2024b; 2024c). The worrying trend

https://doi.org/10.18080/jtde.v12n3.985

of inflation means investing is becoming more important for wealth preservation. Popular investments used by Malaysians include the Employees' Provident Fund (EPF) and the Amanah Saham Bumiputera 2 (ASB2), which had returns of 5.5% and 5%, respectively, in FY2023 (VethaSalam & Ibrahim, 2024; "ASNB declares FY23 dividends", 2023). Despite these investments providing relatively favourable returns, we believe that Deep Reinforcement Learning (DRL) agents capable of ingesting and analysing complex patterns in data can outperform these traditional investing methods by taking advantage of the high volatility present in cryptocurrency prices to yield much higher returns (Tradingview, n.d.) .

In this study, we developed an investing system that trades cryptocurrencies in the daily time frame. The cryptocurrencies selected are Bitcoin (BTC-USD), Ethereum (ETH-USD), BinanceCoin (BNB-USD), DogeCoin (DOGE-USD), Cardano (ADA-USD), Polygon (MATIC-USD), Avalanche (AVAX-USD), and WAVES (WAVES-USD). (While eight cryptocurrencies are listed here, two are later excluded from the final analysis.) The data collected are the historical daily price data of the aforementioned cryptocurrencies retrieved using Yahoo's yFinance API.

The scope of this study includes three DRL algorithms, which are Dueling Deep Q Network (DQN), Advantage Actor Critic (A2C) and Recurrent Proximal Policy Optimization (RPPO). Agents are trained using a sophisticated reward function, which rewards them for profit generation and punishes the agents heavily for losses. Specifically, the proposed reward function multiplies the agents' losses by 3 and profits by 2 in each time step. The impact of losses is intensified to ensure the agents learn risk-averse strategies. The performance metric used to evaluate the agents is the annual return given $1000 during back-testing.

Besides cryptocurrency prices, another source of data is cryptocurrencies' sentiment index, also known as the Fear and Greed index, which is scraped using Alternative.me's API. This API is reliable since Binance has used it to report cryptocurrency sentiments (Binance Square, n.d.). After data ingestion, data preprocessing, and feature engineering, the data is passed as input to the DRL agents. The objective is to enable the agents to leverage both the market data and sentiment analysis for more informed decision-making.

Timeseries prediction is also performed using a four-day window of historical closing prices of the cryptocurrencies. The output of the prediction is used as an additional input feature to the DRL agents to help them learn faster. The timeseries prediction models used were Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU) and Temporal Convolutional Neural Network (TCNN). The objective is to enhance the agents' training efficiency by providing insights into expected future price movements, thereby simplifying their decision-making process during training.

The aim of this study is to develop an intelligent investment system that is able to help users mitigate the impact of inflation and explore a novel investment approach to earn higher returns than those provided by conventional investments. By leveraging Deep Reinforcement Learning and incorporating additional insights, such as sentiment analysis and future price movements, the system aims to make informed trading decisions and learn profitable trading strategies. The ultimate impetus is to help users achieve earlier retirement, financial peace and more control over their lives.

Related Works and Algorithms

Reinforcement learning

The core of this project is based on a type of machine learning called reinforcement learning (RL). This type of learning involves an AI agent that interacts with an environment by observing the state of the environment and taking an action with the goal of maximizing the reward gained by the agent. Some RL algorithms are Value-based, in which they find the value of being in each state (how good a particular state is). Another type of RL algorithm is Policy-based, where they map the observed states to the probability distribution of different actions to take, and then taking the action with the highest probability (Or, 2021). Deep Reinforcement Learning (DRL) is a technique which combines Deep Learning with Reinforcement Learning to solve more complex challenges. These include tasks relying on visual inputs (e.g., gaming with pixels as input).

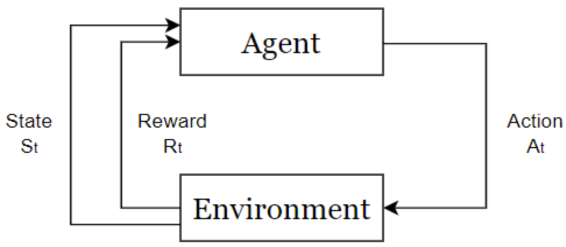


Figure 1. A Reinforcement Learning cycle diagram

Duelling Deep Q Network

In traditional Q learning, a Q table maps state and action pairs to their expected future reward using the Bellman optimality equation:

$$Q^*(s,a) = E\,[R_{t+1} + \gamma\, max_{a'}\, Q_*\,(S_{t+1}, a') |\, S_t = s, A_t = a] \quad (1)$$

where the optimal action-value function $Q^*(s,a)$ is the immediate reward, R_{t+1}, of taking action a, added to the maximum action-value function of all possible actions at the next time step, a', during the future state, S_{t+1}, achieved after taking action a (Sutton & Barto, 2018).

With Deep Q Network (DQN), a neural network is used to approximate the Bellman optimality equation:

$$Q(s, a;\ \theta) \approx Q^*(s, a) \tag{2}$$

This eliminates the need for a large Q table when state-action pairs scale infinitely. The parameter updates are computed with the following equation (Or, 2021):

$$\theta_{k+1} = \theta_k + \ \alpha(r + \gamma max_{a'} Q(s', a';\ \theta_k) \ - Q(s, a;\ \theta_k)\ ) \nabla_{\theta_k} Q(s, a;\ \theta_k) \tag{3}$$

With Duelling Deep Q Network, the computation of how much better an action is compared to all other possible actions in the action space is done with the addition of an Advantage (A) value. The Advantage function providing this value can be represented as:

$$A(s_t,\ a_t) = \ Q_\theta(s_t,\ a_t) - V_v(s_t) \tag{4}$$

The Q value computes the value of taking a specific action at the given state, whereas the V value computes the value of a given state independently of the action taken (Yoon, 2019a). To represent the Q function of the Duelling DQN in a simpler manner (Or, 2021), we can express the function as:

$$Q(s, a) = A(s, a) + V(s) \tag{5}$$

Sornmayura (2019) shows a study using DQN to trade forex such as EUR-USD and USD-JPY with historical price and technical indicators as input features. The author's DQN agent achieved a 43.8% mean annual return on 12 years of EUR-USD historical data and 26.7% for that of USD-JPY, whereas the Commodity Trading Advisor (CTA) only achieved 3.93% on both currencies using the same test sets. However, the author assumed unrealistically in the study that there were no transaction costs involved.

Advantage Actor Critic

Two types of Reinforcement Learning algorithms are Actor methods (Policy-based) and Critic methods (Value-based). DQN is a critic-only method in which we approximate the Bellman equation (1) to estimate the value of each state-action pair, which does not optimize a policy for a given state directly. Actor methods estimate the performance of the gradient with respect to the actor's parameters directly through simulation. This introduces high variance in training, as gradients are calculated independently of past training (Konda & Tsitsiklis, 2003).

The Actor-only method uses the policy gradient for weight updates:

$$\nabla_\theta J(\theta) = \ E_{\ \tau}[\textstyle\sum_{t=0}^{T-1} \nabla_\theta \ log\ \pi\theta\ (a_t|s_t)\ G_t] \tag{6}$$

Equation (6) shows that the previous weights have no effect on gradients (Yoon, 2019b).

A combination of both methods – the Actor-Critic method – is proposed to leverage the strength and eliminate the weakness of both methods. Through simulation, the Critic network will learn the value approximation, which is then used to update the actor's policy parameters. This way, the algorithm will learn a policy with low variance, as the weights are updated using the Q function calculated from the Critic network, which relies on previous updates, as shown in equation (3). The Actor-Critic weight update formula (Yoon, 2019b) is given by:

$$\nabla_\theta J(\theta) = E_\tau[\sum_{t=0}^{T-1} \nabla_\theta \log \pi\theta (a_t|s_t)\, Q_\theta(s_t, a_t)] \tag{7}$$

Advantage Actor Critic just subtracts from the Q function a state-value function $V(s_t)$. This calculates the advantage of taking action, compared to the average action at a given state s. The formula for advantage value (Yoon, 2019b) is:

$$A(s_t, a_t) = r_{t+1} + \gamma\, V_v(s_{t+1}) - V_v(s_t) \tag{8}$$

The Critic network will then be the V function of the Bellman optimality equation, which represents the value of a given state independent of the action taken (Sutton & Barto, 2018). With the Advantage function, equation (7) will be rewritten as:

$$\nabla_\theta J(\theta) = \sum_{t=0}^{T-1} \nabla_\theta \log \pi\theta (a_t|s_t)\, A(s_t, a_t) \tag{9}$$

With these simple mathematical tricks, we have the Advantage Actor Critic (A2C) algorithm (Yoon, 2019b).

The study in Yang *et al.* (2020) shows the results of using A2C in automated stock trading. The authors used the algorithm to trade the Dow Jones 30 Constituent Stocks. The data from January 2009 to September 2015 was used for training, October to December 2015 for validation; and January 2016 to May 2020 was used for back-testing. The state space includes data about the stock prices, stock shares, and remaining balance. The authors used the annual return as well as the Sharpe ratio as an evaluation metric that compares the returns on investment with risks. The ratio divides an investment's excess returns by volatility:

$$\text{Sharpe Ratio} = \frac{R_p - R_f}{\sigma_p} \tag{10}$$

R_p represents the return of portfolio, R_f represents risk-free rate, and the denominator represents the standard deviation of the excess return (Fernando, 2024b). The study in Yang *et al.* (2020) achieved an average Sharpe ratio of 0.1878 and an annual return of 11.4%. According to the authors, A2C is more adaptive to risk, especially when the market is bearish.

Recurrent Proximal Policy Optimization

Proximal Policy Optimization (PPO) is another A2C-based algorithm introduced to ensure that the deviation from previous policy updates is relatively small. The policy objective function at equation (9) is sensitive towards the step size and usually leads to destructively large policy updates (Schulman *et al.*, 2017). PPO ensures the policy updates are conservative, by clipping it between $[1-\varepsilon, 1+\varepsilon]$.

The objective function of PPO can be expressed as:

$$Lt(\theta) = \hat{\mathrm{E}}_t\big(min(\, r_t(\theta)\hat{\mathrm{A}}_t \;, clip(r_t(\theta), 1-\varepsilon, 1+\varepsilon)\hat{\mathrm{A}}_t \,)\big) \tag{11}$$

The ratio of the probability of the new policy and previous policy is denoted by r_t. Intuitively, this ratio is used to calculate the divergence between the policy updates. $\hat{\mathrm{A}}_t$ denotes the estimated advantage at the time step, t. The parameter ε is used to control the clipping, usually set to 0.1 or 0.2. The Minimum function ensures the update does not become too large (OpenAI, 2017).

Due to the sequential nature of financial data, this study will use LSTM layers as input to the PPO algorithm; hence the name Recurrent Proximal Policy Optimization (RPPO).

In addition to A2C, the study in Yang *et al.* (2020) also shows the results of using PPO in trading the Dow Jones 30 Constituent Stocks. With PPO, the authors achieved an average Sharpe ratio of 0.1133 and annual return of 15%, which is the highest among all algorithms tested in the study. According to the authors, PPO has the ability to follow trends and acts well in generating more returns. However, the Sharpe ratio is lower than A2C's, meaning PPO could involve more risks.

To summarize this section, all three proposed algorithms have been used by other studies to perform trading. All methods have been proven profitable. However, our study will improve upon these studies by modelling the trading behaviours of real fund managers. For example, Yang *et al.* (2020) only uses the stock price and trading account information in the state space. Our research extends beyond this by incorporating the process of curating cryptocurrencies, feature engineering of technical indicators, such as Relative Strength Index (RSI), and Moving Average Convergence Divergence (MACD and Bollinger Bands) (Chen, 2021; O'Hara, 2023), as well as understanding market sentiment.

Methodology

Data collection

The primary source of data was the finance API provided by Yahoo Finance. This was used to download the historical price data of the cryptocurrencies. The cryptocurrency data downloaded was for BTC-USD, ETH-USD, BNB-USD, DOGE-USD, ADA-USD, MATIC-USD, AVAX-USD and WAVES-USD. The date of the historical data ranged from 28 January 2018 to 20 December 2021 (most recent data as of the time of this study). This is because the cryptocurrency sentiment (Fear and Greed) index data is only available starting from 1 February 2018, and a 4-day window is used as input. The data was saved into .csv files.

Timeseries prediction

We believe that feeding future price predictions to the DRL agent will help it learn faster, as it will not need to learn predicting future prices on its own. Instead, it can focus on learning trading policies (strategies). Therefore, the second phase of this study involves training a predictive model to predict the future closing prices of the cryptocurrencies.

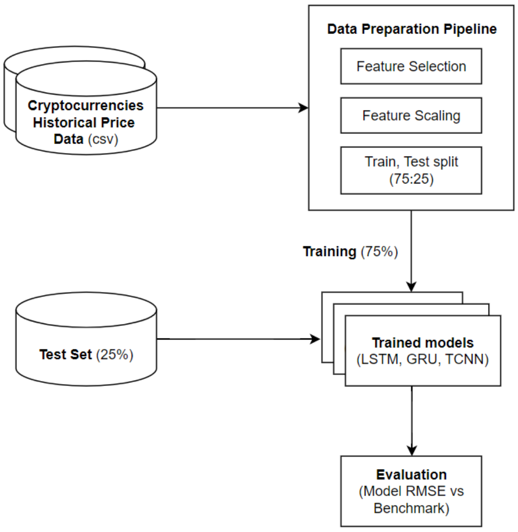


Figure 2. Timeseries prediction flowchart

The data preparation pipeline begins with feature selection. In this phase, we drop the Open, High, and Low columns and only use the Closing Price to do prediction. This is because the other values are highly correlated and are redundant. This will also help prevent overfitting.

The second phase of data preparation is feature scaling. We used the MinMaxScaler from Scikit-learn to scale the values of all cryptocurrencies between 0 and 1. MinMaxScaler largely preserves the relative distance between data points, ensuring the temporal relationships in the data are maintained. This phase ensures the model training is more uniform, as there are no large fluctuations in numbers.

The third phase is train-test split. We split the data using a 75:25 ratio for training and testing, respectively. The predictor variables are the 4-day historical price and the target variable is the actual price of the fifth day. This is the last phase of data preparation.

The next step is to train the prediction models. This study experimented with three different neural network architectures suitable for timeseries applications. They are: Long Short-Term Memory (LSTM); Gated Recurrent Unit (GRU); and Temporal Convolutional Neural Network (TCNN) (Durairaj & Krishna Mohan, 2022; Hamayel & Owda, 2021). To further improve the models' performance, different optimization algorithms, such as Adam, RMSprop, and Stochastic Gradient Descent, were used as well (Fatima, 2020).

Once the models have been trained, they are evaluated with the test set. The prediction is then inverse-scaled for better interpretation of the evaluation results. The metric used is Root Mean Squared Error (RMSE). To make sure the future price prediction is accurate enough to help the DRL agent with prediction, a benchmark RMSE for each cryptocurrency must be achieved, or else the cryptocurrency will be dropped. The benchmark RMSE is finally calculated as:

$$RMSE_{cryptocurrency} < \mu_{closing\ price} * 0.06 \tag{12}$$

The best training results and calculated benchmarks for each cryptocurrency are shown in Table 1.

Table 1. Timeseries prediction benchmarks and results

Cryptocurrency	Best Model	Achieved RMSE	Benchmark RMSE
BTC-USD	TCNN	1688.5353	2052.00
ETH-USD	TCNN	157.5501	146.00
BNB-USD	GRU	22.8944	23.00
DOGE-USD	GRU	0.0087	0.008
ADA-USD	LSTM	0.0722	0.08
MATIC-USD	GRU	0.0975	0.08
AVAX-USD	TCNN	1.5376	3.08
WAVES-USD	GRU	2.0744	0.7

Table 1 shows that only half of the cryptocurrencies exceeded the benchmark given. However, we decided not to drop them yet. To justify this decision, we noted that the cryptocurrencies that achieved an RMSE above the benchmarks only exceeded them by a negligible amount, except WAVES-USD. The second justification is that the benchmark equation (12) is arbitrarily defined. The third justification is that we need more training data for the DRL agents as we only have less than three years of training data for each cryptocurrency.

Web scraping – Cryptocurrency Fear and Greed Index (sentiment analysis)

Market sentiment is another important factor when making trading decisions (Lin *et al.*, 2023). However, it is difficult for us to curate data that accurately represents the market's true sentiment. Limitations, such as data storage, computational power, time constraints and web scraping limitations imposed on different websites, can make it much more difficult to implement sentiment analysis from scratch.

Luckily, the Alternative.me API provides this information. Through the use of the Urllib and BeautifulSoup libraries in Python, we are able to retrieve the cryptocurrency sentiment (Fear and Greed) index calculated daily by Alternative.me.

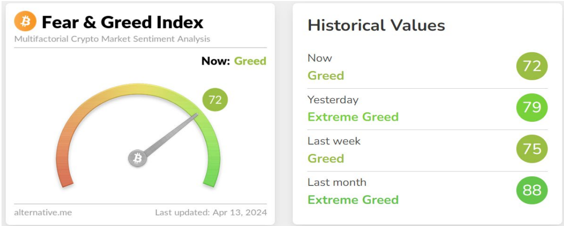


Figure 3. Alternative.me's Fear and Greed Index (Alternative.me, n.d.)

The index ranges from 0 to 100, or "Extreme fear" to "Extreme greed", respectively. The developers of this API ingest data from multiple sources to analyse the emotions and sentiments of Bitcoin and other large cryptocurrencies (Alternative.me, n.d.). The data sources and their weights in the sentiment analysis are shown in Table 2.

Table 2. Fear and Greed index data sources

Data Source	Weight (%)
Volatility	25
Market Momentum/Volume	25
Social Media (Reddit)	15
Surveys	15
Dominance	10
Trends	10

After we have scraped the index, we performed a Pearson's Correlation Coefficient (*r*) analysis to determine whether each of the cryptocurrencies is correlated to this index. Logically speaking, the Fear and Greed Index should be positively correlated with the price of the cryptocurrencies. Therefore, we only select the cryptocurrencies with a positive correlation for DRL training.

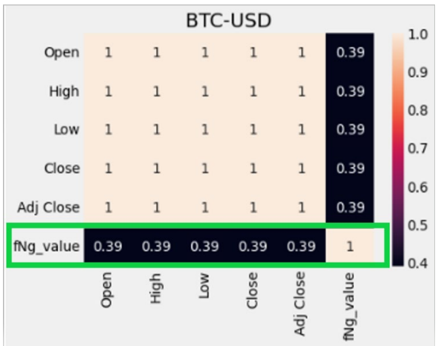


Figure 4. Bitcoin Pearson Correlation Coefficient (*r*) matrix

As expected, Bitcoin has a moderately high positive correlation of 0.39 with the Fear and Greed Index (fNg_value). This means that the index can be used to estimate the future value of Bitcoin. We performed the same analysis for all eight of the cryptocurrencies, using the Adjusted closing price (Table 3).

Table 3. Pearson Correlation Coefficient (*r*) between cryptocurrency prices and Fear and Greed Index

Cryptocurrency	Pearson Correlation Coefficient (r)
BTC-USD	0.39
ETH-USD	0.22
BNB-USD	0.18
DOGE-USD	0.064
ADA-USD	0.19
MATIC-USD	-0.095
AVAX-USD	-0.16
WAVES-USD	0.24

Using the results from Table 3, we decided to drop MATIC and Avalanche (AVAX) from the DRL training set, because they have a negative correlation with the cryptocurrency sentiment (Fear and Greed Index) and likely will not benefit from using it.

Feature engineering

Before going into DRL, there is one extra feature-engineering step to perform, which is computing technical indicators for all cryptocurrencies using their historical price data. The formulas for the technical indicators used are:

Relative Strength Index (RSI):

$$RSI_{step\ one} = 100 - \left[\frac{100}{1+\frac{Average\ gain}{Average\ loss}}\right]$$ (Fernando, 2024a) (13)

Moving Average Convergence Divergence (MACD):

$$MACD = 12\ period\ EMA - 26\ period\ EMA$$ (Dolan, 2024) (14)

where EMA is Exponential Moving Average.

Bollinger Bands:

Upper band = 20-day SMA + (20-day σ x 2) (Lund, 2023) (15)

Lower band = 20-day SMA - (20-day σ x 2) (Lund, 2023) (16)

where SMA is Simple Moving Average and σ is variance.

Deep Reinforcement Learning

The core phase of our study is training the DRL agents to learn the optimal trading policy. In this phase, a custom trading environment class is designed to modularize the processes of storing input action, calculating current and future states, and reward calculations, and a terminator to decide when an agent should stop trading. The environment is a core component of the entire Reinforcement Learning training loop as many training components, such as loss function calculation, input spaces, input observation/states, and training termination, all depend on the calculation of the trading environment.

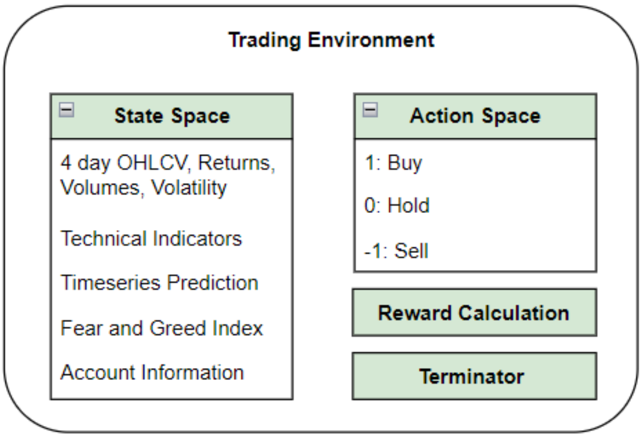


Figure 5. Custom trading environment

The four main components in the trading environment are:

1. State space:
   - 4 days of Open, High, Low, Close, Volume (OHLCV), Returns, Volumes, and Volatility.
   - Technical Indicators: RSI, MACD, Bollinger Bands.
   - Timeseries Prediction.
   - Cryptocurrency sentiment (Fear and Greed) index.
   - Account information: $\frac{Current\ Capital}{Initial\ Capital}, \frac{Running\ Capital}{Current\ Capital}$, Asset Units, Previous Action
2. Action space:
   - 1: Buy
   - 0: Hold
   - -1: Sell
3. Terminator / Stopping case:

   If the agent reaches the end step (last row of the dataset).
   If the agent's capital falls below threshold capital (e.g. Threshold of 50% = $500).
4. Reward Calculation:

   The algorithm to calculate reward in our custom-trading environment is outlined in Figure 6.

Algorithm to calculate reward

Input: C_t: Current capital, C_{t-1}: Previous capital
C_1: Initial capital
A_t: Current action, A_{t-1}: Previous action
AH: Asset held
C_{min}: Minimum capital required to continue trading
$Step_{current}$: Current step
E: End step (last row of data reached)
Output: R: The calculated reward

1 R $\leftarrow C_t - C_{t-1}$
2 // Prevent holding too long
if $A_t = 0$ AND $A_{t-1} = 0$, then R $\leftarrow$R – 5
3 // Prevent selling nothing
if $AH = 0$ AND $A_t = -1$, then R $\leftarrow$R – 50

if $R < 0$, then
    R ←R * 3 // Amplify punishment for losses
else if R >= 0, then
    R ←R * 2 // Amplify reward for gains
end if
// Rewards agent for long term profits.
if E, then R ←R + 10 * ($\frac{C_t}{C_1}$ – 1)
if NOT E AND $C_t < C_{min}$, then
    R ←R – MAX (0.5, 1 – $\frac{Step_{current}}{Step_{end}}$)
    // capital less than threshold
end

Figure 6. Reward calculation pseudocode

After the four components have been set up, all that is left is to train the DRL agents. The algorithms used are the aforementioned Duelling DQN, A2C, and Recurrent PPO. During training, the DRL agents interact with our custom trading environment iteratively, learning optimal trading strategies through trial and error. The agents explore different actions based on their current state, S_t, receive feedback in the form of rewards or penalties, R_t, from the environment, and adjust their behaviour over time to maximize profits. This iterative learning process (Figure 7) enables the agents to adapt their trading policies dynamically, leveraging deep learning to identify patterns and optimize trading strategies in the market environment.

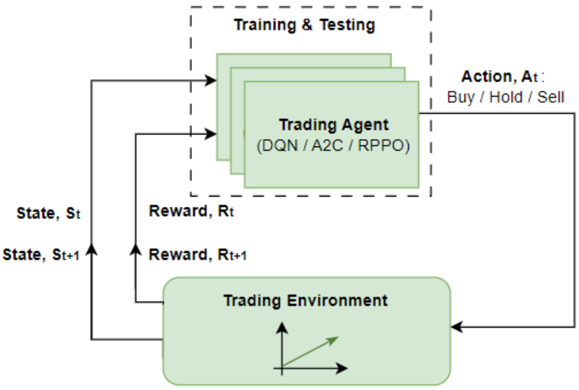


Figure 7. DRL training and testing loop

The models are trained with 60 episodes (equivalent to epochs). In each episode, the models will have traded all 6 cryptocurrencies selected from Table 3 from their earliest date to the latest date in the dataset. GPU acceleration has been enabled for the training via Nvidia's CUDA framework (Pandey *et al.*, 2022). The implementation for DQN and RPPO are in PyTorch, whereas A2C uses Tensorflow.

The important hyperparameters used for the individual models are described in the following subsections.

Duelling Deep Q Network

Table 4. DQN hyperparameters

Hyperparameter	Selected value
Epsilon decay	0.9
Epsilon end	0.1
Gamma	0.9995
Tau	0.001
Learning Rate	0.00005
Memory Length	10000
Memory Threshold	500
Optimizer	Adam
Batch Size	200

On top of Duelling DQN, we also implemented memory replay, which randomly samples training data stored in memory buffers. This removes the sequential dependency of the algorithm as well as makes the algorithm more sample efficient as the experiences are used multiple times (Liu & Zou, 2017). Epsilon is a hyperparameter used to allow the DQN agents to perform exploration (Gimelfarb *et al.*, 2020).

In Reinforcement Learning, exploitation means performing the best action every step based on previous learnings, whereas exploration means the agents are allowed to perform suboptimal actions at the given timestep in hope of escaping local minima and finding a better, previously undiscovered policy (Zangirolami & Borrotti, 2024). For our study, we allow the model to perform more exploration in the beginning, and the epsilon value will decrease to 0.1 near the end of the training, signifying a transition from exploration to exploitation, the justification being that the model should have conducted adequate exploration and found an optimal policy at the end.

Advantage Actor Critic

Table 5. A2C hyperparameters

Hyperparameter	Selected value
Gamma	0.99
Learning Rate	0.0003
Optimizer	Adam

In policy-based learning algorithms like A2C, we have fewer hyperparameters to tune as exploration is built into the algorithm. This is because the algorithm outputs a probability distribution and then samples from that distribution to decide which action to perform. For this algorithm, we decided not to use memory replay as it has slower computation and tends to learn slower compared to PPO.

https://doi.org/10.18080/jtde.v12n3.985

Recurrent Proximal Policy Optimization

Table 6. RPPO hyperparameters

Hyperparameter	Selected value
Gamma	0.99
Learning Rate	0.0003
Batch Size	64
Policy Clip	0.2
Optimizer	Adam

Similar to A2C, we do not need an epsilon value for training PPO as it has exploration built in. However, unlike A2C, we can use memory replay for RPPO, albeit with a smaller batch size of 64 compared to DQN, as LSTM layers require more memory. The policy clip, which has been explained in equation (11), will be set to 0.2 to limit the policy updates.

The training and testing have been performed on Nvidia's GTX1050 GPU and Intel's Core i5-8300H CPU with a RAM of 8GB.

Table 7. Training and testing time

Agent	Time taken (HH:MM)
DQN	00:54
A2C	05:37
RPPO	21:15

Table 7 shows that a value-based model such as DQN is much faster compared to A2C and RPPO, which requires two neural networks to be trained. Furthermore, RPPO uses LSTM cells, which further slows down the training and inference.

Results and Discussion

Training Rewards

To determine the learning behaviours of the DRL agents during training, we can plot the training scores gained over the 60 training episodes.

Figures 8–10 show the improvement in training rewards over 60 training episodes. The scale of the rewards in the y-axis is unimportant as there are more punishments in the reward calculation, as shown in Figure 6. Instead, the main purpose of these plots is to show the training of the agents.

DQN has been proven to be effective at learning trading policies. We can see that DQN training rewards fluctuate more during the early episodes. However, when DQN reaches the later episodes (as denoted in the green box of Figure 8), we can see that the training becomes more stable, trending slightly upwards and reaching maximum at episode 59. The model weights at the peak reward will be saved for evaluation.

https://doi.org/10.18080/jtde.v12n3.985

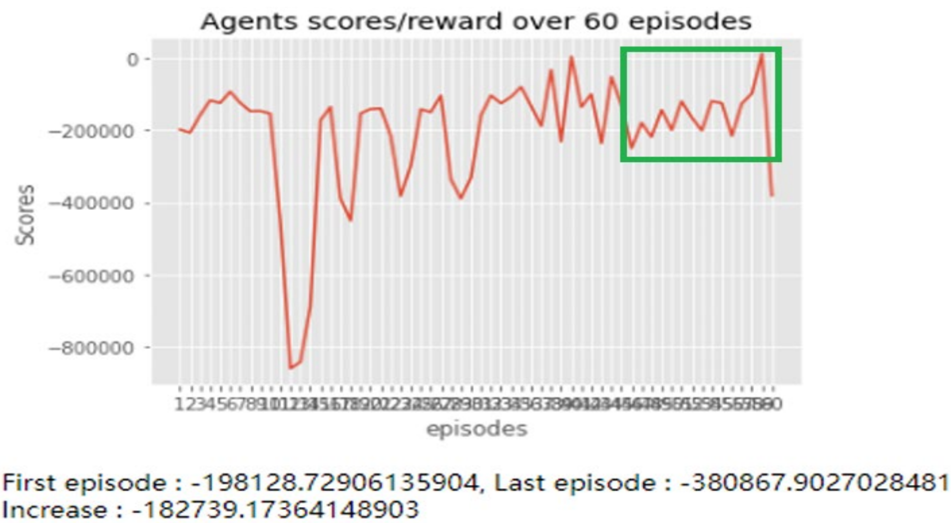


Figure 8. DQN training rewards

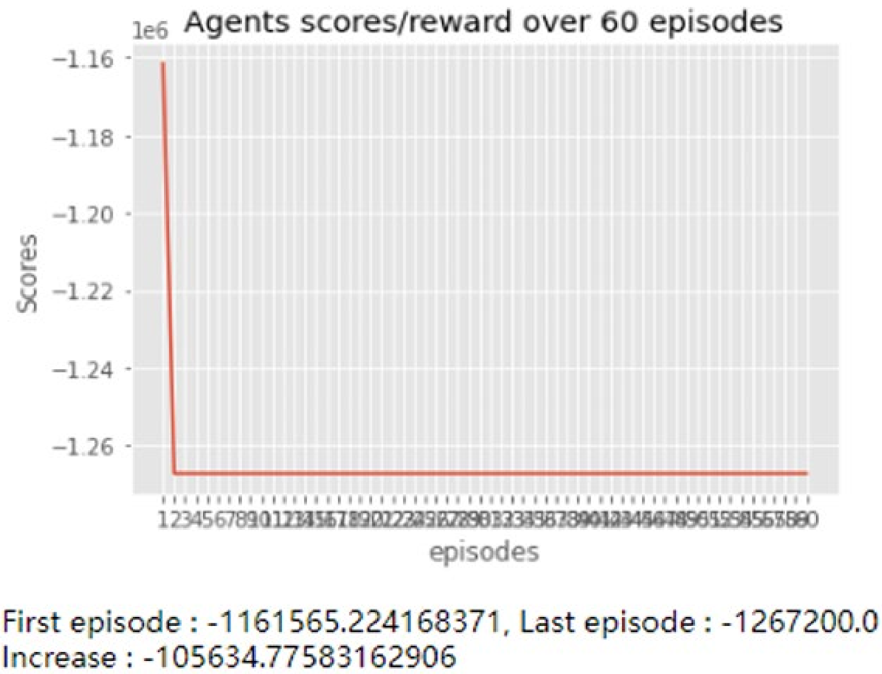


Figure 9. A2C training rewards

In contrast, A2C shows no sign of improvement since episode 1 (Figure 9). Several reasons could have led to this phenomenon, such as the lack of experience replay or the high variance in the policy updates.

The RPPO agent shows that the model was trying to learn an optimal policy (Figure 10); however, it did not show much sign of improvement as it fluctuates around the mean. Comparing the results of DQN and RPPO, it seems that DQN performed better during training, reaching positive rewards at the end of the training, whereas RPPO stayed below 0 the whole time.

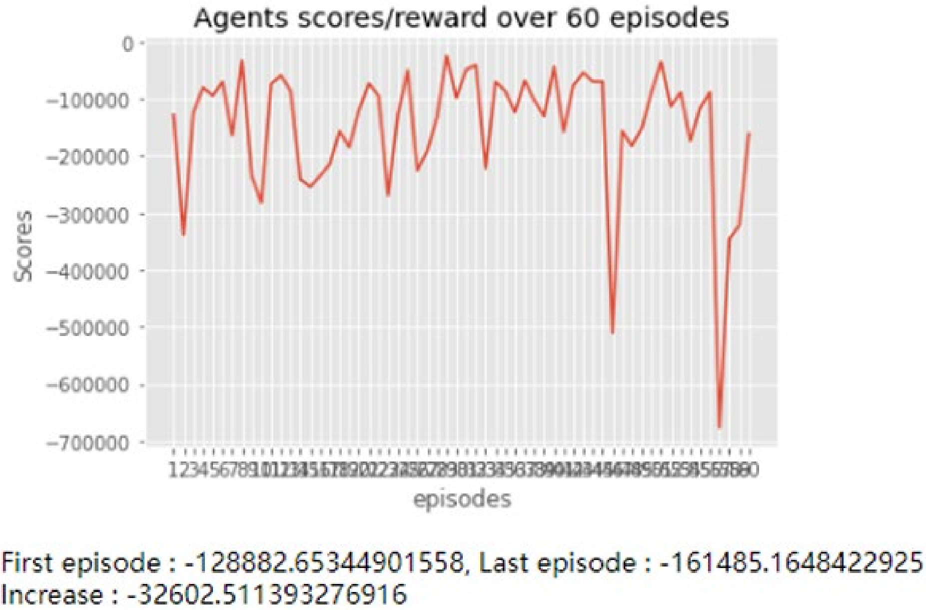


Figure 10. RPPO training rewards

Annual Return

As discussed before, the performance metric used to evaluate all the models was the annual return of trading $1000 on the last year of each cryptocurrency (test set). We also compared them to a random agent that took actions randomly. The random agent was expected to make a loss (due to transaction costs) and acted as a baseline to determine if DRL is effective at learning trading.

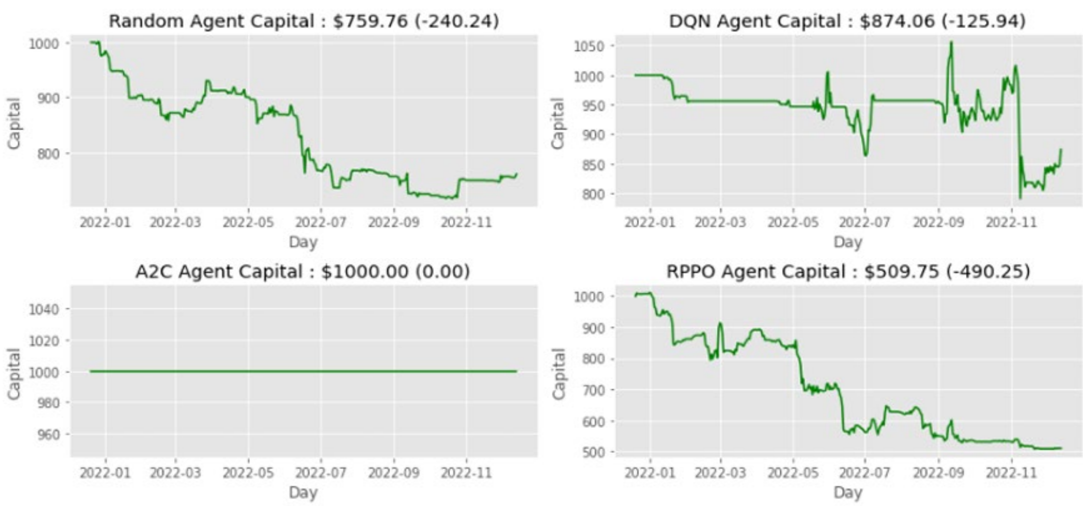


Figure 11. BTC-USD annual return

 https://doi.org/10.18080/jtde.v12n3.985

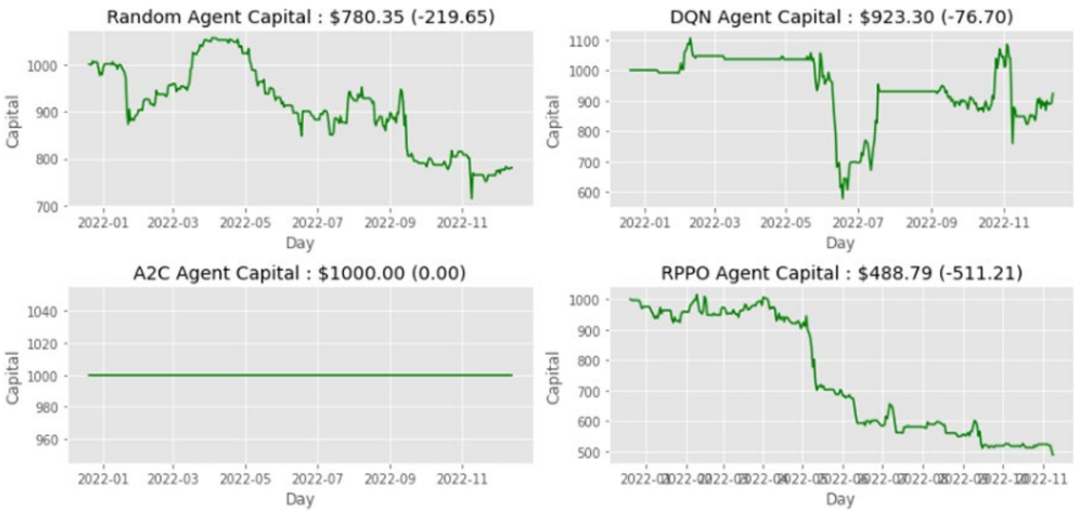


Figure 12. ETH-USD annual return

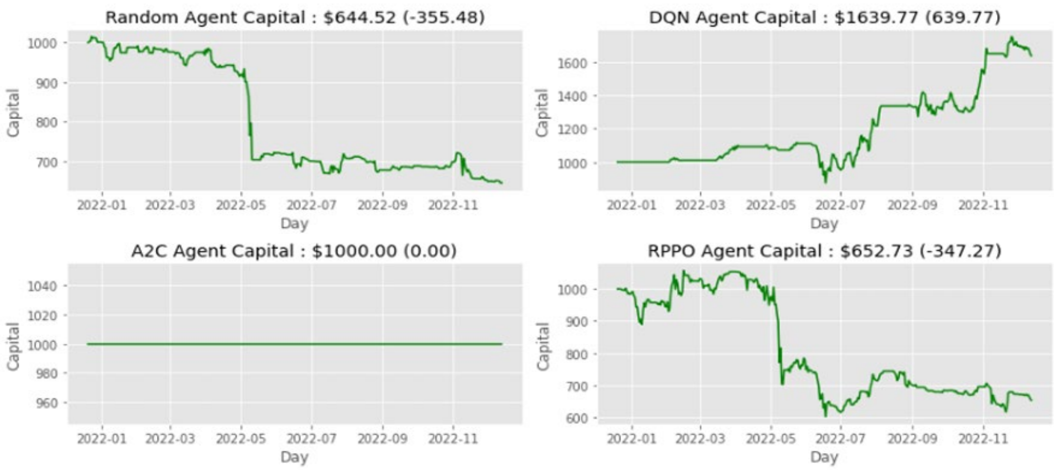


Figure 13. BNB-USD annual return

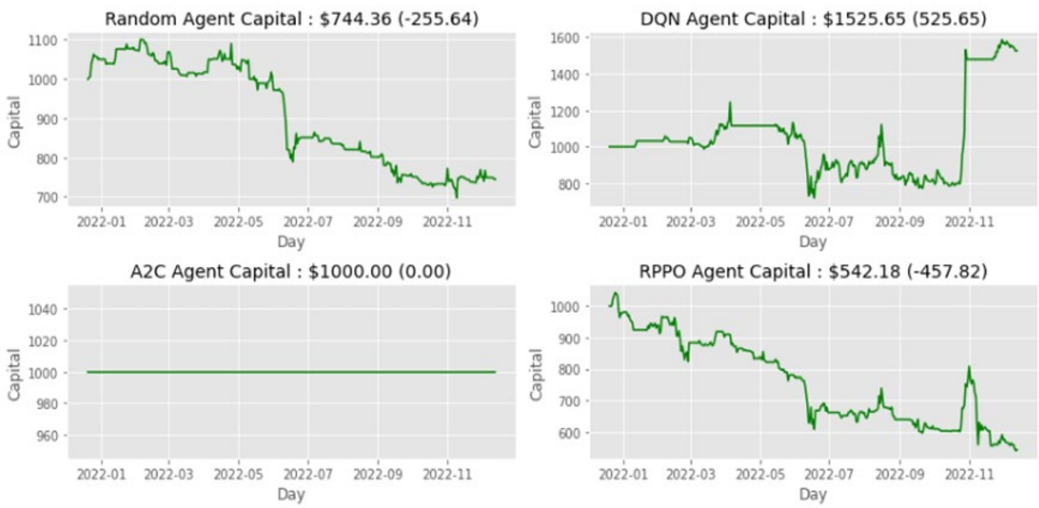


Figure 14. DOGE-USD annual return

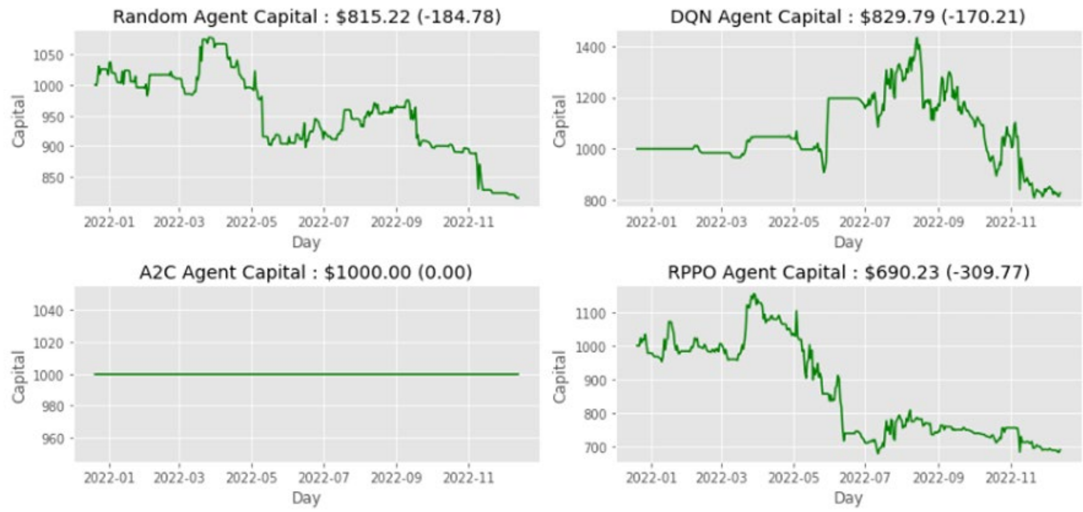


Figure 15. ADA-USD annual return

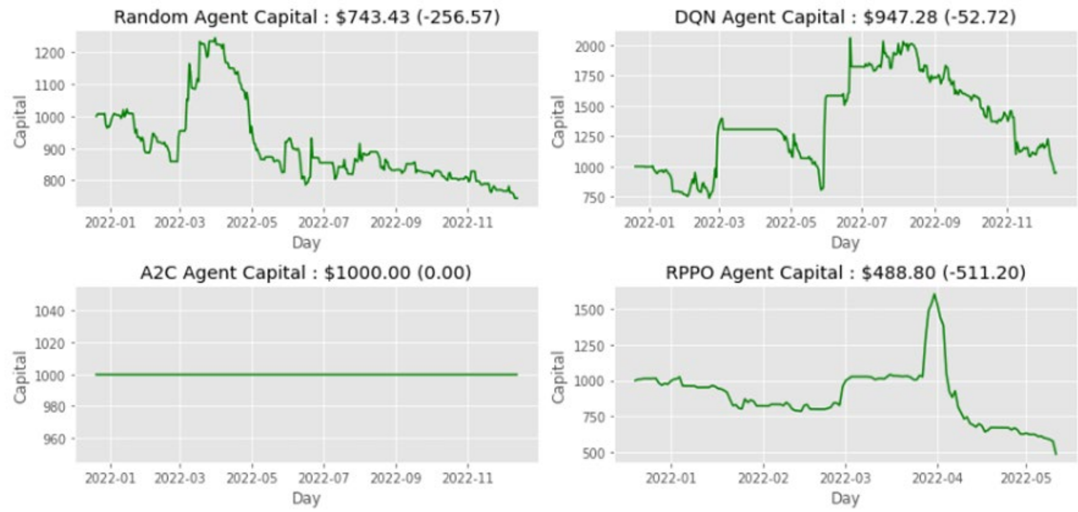


Figure 16. WAVES-USD annual return

Table 8. Summary of annual return achieved

Cryptocurrency	Best Agent	Random Agent's profit	Profit gained	Annual Return
BTC-USD	DQN	-$240.24	-$125.94	-12.59%
ETH-USD	DQN	-$219.65	-$76.70	-7.67%
BNB-USD	DQN	-$355.48	$639.77	63.98%
DOGE-USD	DQN	-$255.64	$525.65	52.57%
ADA-USD	DQN	-$184.78	-$170.21	-17.02%
WAVES-USD	DQN	-$256.57	-$52.72	-5.27%

Table 8 shows that the DQN is the best performing agent for all cryptocurrencies. It also shows that it is better than the random agent, which means it was able to learn some profitable trading policies. The agent profited at trading BinanceCoin and DogeCoin but made a loss when trading the other cryptocurrencies. In BNB-USD, the agent was able to earn the highest return of 63.98%. More importantly, it achieved this return through steady growth of capital rather than winning a few lucky trades, as shown in Figure 13. The average annual return across all cryptocurrencies was 12.33%, which means the agent is profitable overall. If we only trade the cryptocurrencies that the agent is profitable in, we can get an average annual return of 58.27%. This is much better than our initial expectation.

To better visualize the trading decisions made by the DQN agent, we can plot the trading decisions taken by the agent on the most profitable and least profitable cryptocurrencies, which are BNB-USD (Figure 17) and ADA-USD (Figure 18), respectively.

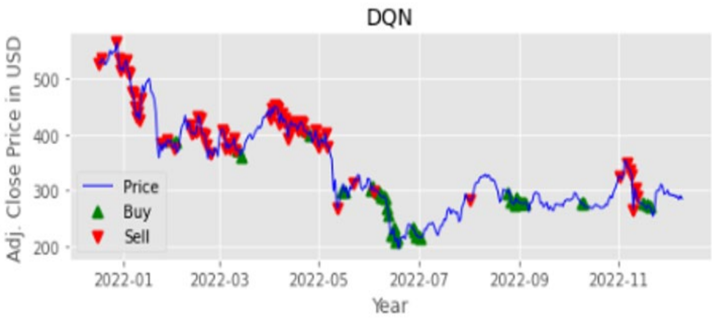


Figure 17. DQN trading decisions on BNB-USD

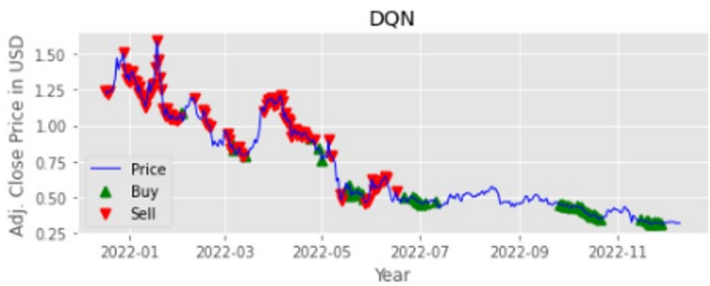


Figure 18. DQN trading decisions on ADA-USD

By analysing Figure 17 and Figure 18, we note that the model was able to capture the trading signals "buy low" and "sell high". The reason ADA-USD appears to have made a loss in Table 8 is because the agent bought low around October 2022. This is most likely due to the anticipation of a retracement. However, the dataset cuts off before we can prove that theory. Furthermore, if we disregard the data before September 2022, we can see that the agent actually profits, as shown in Figure 15.

Conclusion

Based on the results and our analysis, we can conclude that with Deep Reinforcement Learning combined with profound feature engineering, environment setup, and reward design, a profitable trading policy is attainable. The most impressive result is that our best agent, DQN, did not once hit the terminal state (losing more than 50% of the initial capital) for all cryptocurrencies. Cryptocurrencies are volatile in nature and possess a lot of risk. However, our agent was able to control its trading risk and maintain a capital of above $500. The driving factor of this behaviour is attributable to the detailed design of the reward calculation algorithm, which we have experimented with and tuned.

https://doi.org/10.18080/jtde.v12n3.985

Several technical insights can be gained from this study. Firstly, careful consideration of the reward calculation is an important factor in the success of Deep Reinforcement Learning. Unlike traditional supervised learning, we can creatively design our own reward calculation to cater for our own preferences and needs. Secondly, the utilization of the memory replay in reinforcement learning has shown good results. DQN and RPPO both use replay and were able to learn some trading policies, despite RPPO being ultimately unprofitable. On the other hand, A2C was not able to learn anything, as shown in Figure 9.

This study highlighted the feasibility and potential of Deep Reinforcement Learning in the domain of investing. This provides users the ability to invest without intense study and constant monitoring of the market, since this can be done automatically by our proposed system. Our DQN model achieved a return of 63.98% on BinanceCoin and 12.33% overall, which is able to help investors beat the rising inflation rate of 1.8% (as of February 2024).

Future work can include using the Sharpe ratio for evaluation metrics; carrying out more hyperparameter tuning; using better compute, which allows larger neural network architectures; experimenting with different technical indicators; and trading different asset classes such as stocks. Technical indicators other than the ones used in this work (RSI, MACD, Bollinger Bands) may also be experimented with to determine their effectiveness. Finally, more exploration of the timeseries prediction can be carried out, such as experimenting with different algorithms, features and scaling techniques.

References

Alternative.me. (n.d.). Crypto Fear & Greed Index. *Alternative.me*. Retrieved April 10, 2024, from https://alternative.me/crypto/fear-and-greed-index/

"ASNB declares FY23 dividends". (2023, March 31). *The Star*. https://www.thestar.com.my/business/business-news/2023/03/31/asnb-declares-fy23-dividends

Binance Square. (n.d.). Crypto Fear & Greed Index. *Binance*. Retrieved April 10, 2024, from https://www.binance.com/en/square/fear-and-greed-index

Chen, J. (2021, September 29). Technical Indicator: Definition, Analyst Uses, Types and Examples. *Investopedia*. https://www.investopedia.com/terms/t/technicalindicator.asp

Dolan, B. (2024, March 8). What Is MACD? *Investopedia*. https://www.investopedia.com/terms/m/macd.asp

DOSM. (2024a, January). Consumer Price Index December 2023. Department of Statistics Malaysia. https://www.dosm.gov.my/portal-main/release-content/consumer-price-index-december-2023

DOSM. (2024b, March). Consumer Price Index February 2024. Department of Statistics Malaysia. https://www.dosm.gov.my/portal-main/release-content/consumer-price-index-february-2021

https://doi.org/10.18080/jtde.v12n3.985

DOSM. (2024c, February). Consumer Price Index January 2024. Department of Statistics Malaysia. https://www.dosm.gov.my/portal-main/release-content/consumer-price-index-january-2024

Durairaj, M., & Krishna Mohan, B. H. (2022). A convolutional neural network based approach for financial time series prediction. *Neural Computing and Applications, 34*, 13319–13337. https://doi.org/10.1007/s00521-022-07143-2

Fatima, N. (2020). Enhancing Performance of a Deep Neural Network: A Comparative Analysis of Optimization Algorithms. *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal*, *9*(2), 79–90. https://doi.org/10.14201/ADCAIJ2020927990

Fernando, J. (2024a, April 10). Relative Strength Index (RSI) Indicator Explained With Formula. *Investopedia*. https://www.investopedia.com/terms/r/rsi.asp

Fernando, J. (2024b, January 30). Sharpe Ratio: Definition, Formula, and Examples. *Investopedia*. https://www.investopedia.com/terms/s/sharperatio.asp

Gimelfarb, M., Sanner, S., & Lee, C. G. (2020). Epsilon-BMC: A Bayesian Ensemble Approach to Epsilon-Greedy Exploration in Model-Free Reinforcement Learning. Proceedings of the 35th Uncertainty in Artificial Intelligence Conference, PMLR, *115*, 476–485. https://doi.org/10.48550/arXiv.2007.00869

Hamayel, M. J., & Owda, A. Y. (2021). A Novel Cryptocurrency Price Prediction Model Using GRU, LSTM and bi-LSTM Machine Learning Algorithms. *AI*, *2*(4), 477–496. https://doi.org/10.3390/ai2040030

Konda, V. R., & Tsitsiklis, J. N. (2003). Actor-Critic Algorithms. *Neural Information Processing Systems, 42*(4), 1143–1166.

Lin, M., Meng, Y., & Zhu, H. (2023). How connected is the crypto market to risk to investor sentiment? *Finance Research Letters, 56*, 104177. https://doi.org/10.1016/j.frl.2023.104177

Liu, R., & Zou, J. (2017). The Effects of Memory Replay in Reinforcement Learning. *arXiv, 1*(06574). https://doi.org/10.48550/arXiv.1710.06574

Lund, B. (2023, July 11). Bollinger Bands: A powerful technical tool for traders. *Britannica Money*. https://www.britannica.com/money/bollinger-bands-indicator

OpenAI. (2017, July 20). Proximal Policy Optimization. https://openai.com/research/openai-baselines-ppo

Or, B. (2021, January 30). Value-based Methods in Deep Reinforcement Learning. *Towards Data Science*. https://towardsdatascience.com/value-based-methods-in-deep-reinforcement-learning-d40ca1086e1

O'Hara, N. (2023, August 31). The Multiple Strategies of Hedge Funds. *Investopedia*. https://www.investopedia.com/articles/investing/111313/multiple-strategies-hedge-funds.asp

Pandey, M., Fernandez, M., Gentile, F., Isayev, O., Trospha, A., & Stern, A. C. (2022). The transformational role of GPU computing and deep learning in drug discovery. *Nature Machine Intelligence, 4*, 211–221. http://dx.doi.org/10.1038/s42256-022-00463-x

https://doi.org/10.18080/jtde.v12n3.985

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal Policy Optimization Algorithms. *arXiv, 2*(06347). https://doi.org/10.48550/arXiv.1707.06347

Sornmayura, S. (2019). Robust FOREX Trading with Deep Q Network (DQN). *ABAC Journal, 39*(1), 15–33. https://core.ac.uk/download/pdf/233618241.pdf

Sutton, R. S., & Barto, A. B. (2018). *Reinforcement Learning: An introduction* (2nd ed.). London, England: The MIT Press.

Tradingview. (n.d.). BTCUSD chart. *Tradingview*. Retrieved April 10, 2024, from https://www.tradingview.com/chart/cZxzp0Jc/?symbol=BITSTAMP%3ABTCUSD

VethaSalam, R., & Ibrahim, J. (2024, March 3). EPF declares a 5.5% dividend for conventional savings for 2023. *The Star*. https://www.thestar.com.my/news/nation/2024/03/03/epf-declares-55-dividend-for-conventional-savings-for-2023

Yang, H., Lim, X. Y., Zhong, S., & Walid, A. (2020). Deep Reinforcement Learning for Automated Stock Trading: An Ensemble Strategy. In ACM International Conference on AI in Finance, *31*, 1–8. https://dx.doi.org/10.2139/ssrn.3690996

Yoon, C. (2019a, October 20). Dueling Deep Q Networks. *Towards Data Science*. https://towardsdatascience.com/dueling-deep-q-networks-81ffab672751

Yoon, C. (2019b, February 6). Understanding Actor Critic Methods and A2C. *Towards Data Science*. https://towardsdatascience.com/understanding-actor-critic-methods-931b97b6df3f

Zangirolami, V. & Borrotti, M. (2024). Dealing with uncertainty: Balancing exploration and exploitation in deep recurrent reinforcement learning. *Knowledge-Based Systems, 293*, 111663. https://doi.org/10.1016/j.knosys.2024.111663

Developing a Holistic Framework for Assessing the Development of the Digital Economy

A Systematic Review of Key Dimensions and Indicators

Zhong Ni
Faculty of Accountancy and Management, Universiti Tunku Abdul Rahman, Malaysia
Suzhou City University, China

Cheng Ming Yu
Faculty of Accountancy and Management, Universiti Tunku Abdul Rahman

Tey Sheik Kyin
Faculty of Accountancy and Management, Universiti Tunku Abdul Rahman

Fok Kuk Fai
Faculty of Accountancy and Management, Universiti Tunku Abdul Rahman

Abstract: Rapid advancements in information and communication technologies (ICT) have accelerated the expansion of the digital economy (DE). However, the lack of a universally accepted definition has hindered a systematic assessment of the progress of the DE. A robust framework to evaluate various aspects of DE activities is crucial to fully capture the potential of the DE. This study conducts a systematic review of academic literature to identify and analyse various dimensions of the DE, aiming to propose a framework to assess the development of the DE. A comprehensive analysis of 161 articles from the Scopus database (2010–2024) identifies four core dimensions, namely digital infrastructure, digital industrialisation, industrial digitisation and digital innovation as the primary focus of existing DE research. These findings lay the groundwork for developing the framework to assess the development of the DE.

Keywords: digital economy/DE; framework; VOSviewer; systematic review

Introduction

The global economic landscape has undergone a profound transformation in recent decades due to rapid technological advancements and globalisation. This transformation has gained

https://doi.org/10.18080/jtde.v12n3.988

even more significance in the era of the Fourth Industrial Revolution, and has been further heightened by the recent COVID-19 pandemic. The pervasive influence of new information technologies, such as 5G, artificial intelligence and the Internet of Things, is rapidly reshaping all aspects of economic and social life (Zhang *et al.*, 2021). In the current context of slow global economic growth, the continual and rapid expansion of the digital economy (DE) has played a vital role in alleviating downward economic pressures and significantly contributing to driving global economic recovery (Jiang, 2020). As a result, the DE is flourishing and emerging as a crucial driver of global economic and social development.

A robust DE is widely recognised as a key to national success. This realisation has spurred both developed and developing nations to prioritise the development of digital sectors within their respective economies. Developed nations are taking the lead by acknowledging the importance of the DE and planning strategies for its growth. For instance, the United States Department of Commerce plays a pivotal role in driving the country's DE forward. To maintain its position as a global leader in information technology (IT) innovation and digital outcomes, the United States has implemented a range of DE policies and initiatives. These include the Digital Economy Agenda (2015) (Davidson, 2015), Enabling Growth and Innovation in the Digital Economy (2016), the United States' National Cyber Strategy (2018), and the United States' Grand Strategy for the Global Digital Economy (2021) (Atkinson, 2021). The United Kingdom has also elevated the DE to a national strategy; commencing with the Digital Economy Strategy (2015–2018), followed by the United Kingdom Digital Strategy (2017), and culminating in the National Data Strategy (2020). These strategic blueprints represent successive endeavours towards the realisation of a world-leading DE (Department for Digital, Culture, Media & Sport, 2022).

Following in the footsteps of developed nations, developing countries have likewise formulated strategies to propel the advancement of their DEs. In 2015, India introduced the "Digital India" initiative. This strategy aims to foster the digital transformation of India's economy and society, aligning with global trends towards digitalisation and technological innovation (Tripathi & Dungarwal, 2020). In 2018, China formally unveiled the Strategic Outline for the Development of the DE (CAICT, 2023), outlining a comprehensive strategic blueprint for the systematic advancement of digital infrastructure and services to foster the growth of the DE. Subsequently, in 2023, China released the Overall Layout Plan for the Construction of Digital China, which provided a comprehensive strategic blueprint for the overarching development of digital China (The State Council of the People's Republic of China, 2023).

To effectively integrate the DE into national development strategies, it is imperative to be able to assess the current state of DE development. It is evident that the DE has emerged as a focal point for countries worldwide. The need to evaluate the development of the DE has gained significant prominence as more nations incorporate it into their economic development strategies (Millar & Grant, 2019).

The rapid growth of the DE highlights a critical research gap: the lack of a systematic and holistic framework for assessing its development. This gap is evident in the lack of a unified definition and evaluation system to measure the progress of the DE. Some scholars resort to using single and narrow indicators (Yang, 2023), failing to capture the multidimensional nature of the DE adequately. Even those who create multidimensional indicator systems often limit their focus to aspects like information and communication technology (ICT) and the Internet, overlooking a comprehensive understanding of the DE. For instance, Rehman & Nunziante (2023) constructed a DE measurement index based on one indicator each from the dimensions of "e-government, e-commerce and internet", which oversimplifies the complexity of the DE. Additionally, some scholars exhibit overlapping attributes within their chosen indicators, as seen in the work of Zhang *et al.* (2023), who incorporated metrics such as the number of Internet users and Internet penetration rate into their index system simultaneously.

To address this critical research gap, it is essential to conduct a comprehensive review and synthesis of existing studies to identify the key dimensions or elements of the DE and indicators to assess the progress of DE development. Through this review, this study aims to achieve the following objectives:

1. Examine the evolving trends in academic publications related to the DE, focusing on key areas, terminology and core activities of the DE.
2. Identify fundamental elements, dimensions and indicators used in past studies to assess DE development.
3. Propose a framework for assessing the development of the DE.

Literature Review

As definitions are influenced by the times and trends from which they emerge (Bukht & Heeks, 2017), the concept of the DE has transformed in parallel with the evolution of its underlying technologies. The emergence and growth of the DE can be traced back to the advent of the Internet. The next section will first look at the definition of DE to uncover the key dimensions or elements of the DE, as well as tracing the evolution of the assessment methodologies over time.

https://doi.org/10.18080/jtde.v12n3.988

The definition of DE

Bukht & Heeks (2017) provided a comprehensive definition of the DE and proposed a three-tier framework for its progression. Firstly, at the core of the DE is the IT/ICT sector. In the second tier, the DE, in a narrow sense, is defined as an economy where the primary output is derived from digital goods or services based on digital technologies. The third tier, characterising the DE in a broad context refers to the use of ICT in all sectors of the economy, denoted as the "digitalised economy". Scholars have extensively drawn upon this framework since its introduction (Williams, 2021).

From the core level, the DE focuses exclusively on the ICT industry and digital technology. As a pioneer in the development of the DE, Don Tapscott offered a comprehensive exploration of the transformative effects of the computer and Internet revolution on business activities (Tapscott & Caston, 1994). This work aligns with the United States Department of Commerce's summary of the characteristics of the DE, which includes the Internet as infrastructure, IT as a leading technology, the information industry as a leading and supporting industry, and e-commerce as an engine of economic growth (Henry *et al.*, 1999). Moulton (2000) further defined DE to encompass e-commerce, IT, infrastructure and information transmission, communication, and computer industries. Due to this emphasis on Internet connectivity and IT, the DE is sometimes referred to as the "internet economy" (Irkinovich, 2022).

A narrow view of the DE, supported by scholars like Knickrehm *et al.* (2016), focuses solely on economic activities reliant on the ICT sector. In contrast, a broader perspective, as outlined by the G20 (2016), encompasses a wider range of economic activities. This broader definition includes the use of digitised information and knowledge as the key production factor, modern information networks as the primary activity space, and the effective use of ICT as a driver for efficiency and economic restructuring.

A broader definition that underscores the evolutionary nature of the DE is warranted (Hanna, 2020). The China Academy of Information and Communications Technology (CAICT) extended the G20's definition and defined DE as "a new form of economy that takes digitalised knowledge and information as the key production elements, digital technology innovation as the core driving force, and modern information networks as important carriers" (CAICT, 2017). The DE is a distinct and novel economic paradigm that is underpinned by modern IT, infrastructure and data (Ma *et al.*, 2022), succeeding the agricultural and industrial economies (Lange *et al.*, 2020).

Evidently, the DE has evolved in tandem with advancements in IT. Initially confined to the ICT sector, the scope gradually expanded to include economic activities driven by digital technology. Presently, it is widely recognised as a new economic form rooted in digital

https://doi.org/10.18080/jtde.v12n3.988

technology (Shahbaz *et al.*, 2022). With the rapid advancement of IT and the DE, the conventional frameworks that define the DE solely through core and narrow perspectives are becoming inadequate. The DE's influence extends beyond singular industry sectors, impacting all facets of the economy and daily life (Wang *et al.*, 2022). As its scope and content continue to evolve, adopting a comprehensive and broader definition is essential for a thorough understanding and representation of the DE. This broader perspective allows for a better comprehension of its developmental trajectory and facilitates the formulation of more comprehensive and effective policies.

The measurement for assessing DE development

As there is no uniform and comprehensive definition of the DE, there are various methods to assess the development of the DE. Creating a DE index utilising evaluation indicators serves not only to monitor the developmental trajectory of the DE but also to identify key enablers fuelling its progress. This approach offers a comprehensive and timely assessment of the DE's evolution, revealing dynamic features and trends. These insights inform policymaking, business strategies and research endeavours aimed at facilitating digital transformation.

For instance, the Digital Economy and Society Index (DESI) has been published by the European Commission (2022) since 2014 to evaluate the development of digitalisation in the European Union (EU) member states. Based on a comprehensive set of over 30 indicators, DESI measures various aspects of the DE and society in five areas, including connectivity, use of the Internet, human capital, integration of digital technology, and digital public services (European Commission, 2022). It has gained significant popularity among academics and policymakers (Bruno *et al.*, 2023). It should be noted that in addition to DESI, there is the I-DESI (the International DESI). It is an extended and advanced version of DESI and extends its scope beyond the EU to encompass a comparative analysis of DE performance among EU member states and 18 other countries globally (Nagy & Somosi, 2022).

The Network Readiness Index (NRI), initially introduced by the World Economic Forum (WEF) in 2001, acts as a quantitative tool to evaluate a country's preparedness in utilising ICT for active engagement in the DE. In its 2023 edition, the NRI retains its four-pillar framework for structural assessment, encompassing technology, people, governance and impact (Portulans Institute, 2023). This framework incorporates a total of 58 indicators, demonstrating a meticulous approach aimed at evaluating a nation's readiness for DE participation.

Apart from using established indicators, many scholars create their own indicator systems to assess various aspects of the DE and its development. For instance, some scholars assess the

https://doi.org/10.18080/jtde.v12n3.988

DE from two dimensions, such as digital inclusive finance and Internet development (Cheng *et al.*, 2023). Others expand their models to include three dimensions, such as:

- informatisation, Internet and digital transaction (Zhang *et al.*, 2023)
- infrastructure, industrial scale and spillover value (Pan *et al.*, 2022)
- digital infrastructure, application of digital technology and development of the digital industry (Yang & He, 2022)
- digital infrastructure, digital industrialisation and industrial digitisation (Li & Zhao, 2023).

The framework can also go beyond three dimensions, incorporating elements such as digital infrastructure, digital application, digital innovation and digital benefits (Su *et al.*, 2023).

Multiple dimensions illustrate the multifaceted nature of DE measurement. However, to fully understand the key dimensions or scope of the DE, and the evolvement of the assessment methods, a comprehensive review of the existing literature is necessary. A thorough examination of the literature would allow us to systematically analyse the various applications of measurement approaches, their effectiveness in capturing different aspects of DE development, and any identified challenges. This knowledge is critical for constructing an index that effectively captures the development of the DE.

Methodology

This study employs a combined Systematic Literature Review (SLR) and Bibliometric Analysis (BA) as well as Content Analysis (CA) approach to comprehensively analyse the existing literature to identify the key indicators of DE development.

First, this study employs SLR technique to systematically screen and select relevant research papers. SLR provides a structured method for identifying, selecting and synthesising relevant research, ensuring a rigorous and transparent review process (Khan & Abdullah, 2021).

Subsequently, the selected papers will undergo BA and CA. BA summarises extensive bibliometric data to delineate the structure of knowledge and identify emerging trends within a research topic (Donthu *et al.*, 2021). It provides a quantitative assessment to examine publication trends and patterns related to DE development (Anthopoulos & Tzimos, 2021). CA is a widely utilised analytical method in qualitative research. This method involves gathering data that fits within specified concepts and themes, followed by interpreting this data in a manner that is comprehensible to the reader (Aktoprak & Hursen, 2022). And it provides in-depth analysis of the content to identify commonly used dimensions and indicators for measuring DE development. By integrating these three methods, this research aims to advance the understanding of the DE beyond the traditional SLR studies (Pulsiri & Vatananan-Thesenvitz, 2018).

Data collection

Publication data were collected from Scopus database. Scopus is widely recognised as one of the leading bibliometric databases (Baquee *et al.*, 2023) due to its extensive coverage and user-friendly interface. It is often regarded as one of the largest curated databases, encompassing scientific journals, books, conference proceedings and more (Singh *et al.*, 2021). These materials are selected through a rigorous process of content curation and ongoing evaluation. These features ensure a comprehensive search for relevant research on DE measurement for this study.

Literature screening

Articles published between 2010 and 12 March 2024 were retrieved and screened followed the steps outlined in Figure 1. The 2010–2024 timeframe was chosen to capture the period of burgeoning research interest in DE. This study focused on terms associated with the DE and explored methods for measuring its development. Specifically, this study looked for titles, keywords or abstracts containing "digital economy" or "internet economy" combined with "indicator," "measure," or "index". The search keys are TITLE-ABS-KEY (("digital economy" or "internet economy") and ("indicator" or "measure" or "index")).

Following the initial search, a total of 1,826 potential results were identified. Subsequently, the scope was refined to include only English-language articles, focusing specifically on journal papers and conference papers. Considering the contemporary nature of the DE as a subject of study, conference papers were deemed relevant for inclusion. This filtering process yielded 1,500 papers for further analysis. To ensure the study's focus on DE measurement practices, the search was limited to specific disciplines. The inclusion of "Social Sciences," "Business, Management and Accounting," and "Economics, Econometrics and Finance" yielded the most relevant research directly related to DE measurement and analysis.

A thorough examination of titles and abstracts led to the exclusion of 452 articles that did not meet the selection criteria. This resulted in a total of 220 articles for further analysis. A full-text screening process further excluded 59 articles that were either not in English, lacked full-text availability, or did not directly address the designated topic. Ultimately, a refined selection of 161 articles was obtained for comprehensive review and synthesis.

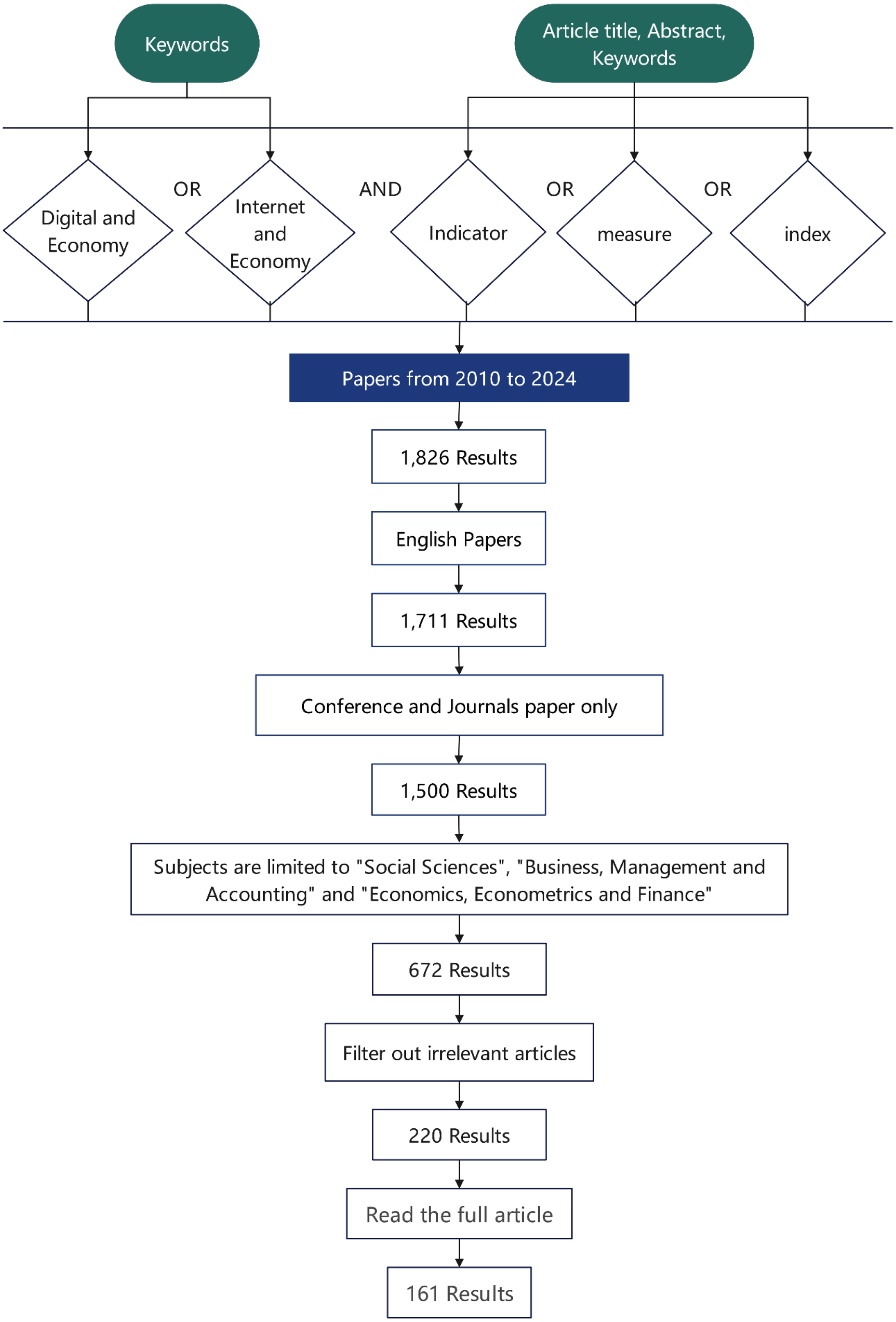


Figure 1. Flowchart of data collection and screening process in SLR Framework

Data analysis

To gain deeper insights into the research related to the DE and identify key indicators of DE development, bibliometric and content analyses were conducted on the selected articles.

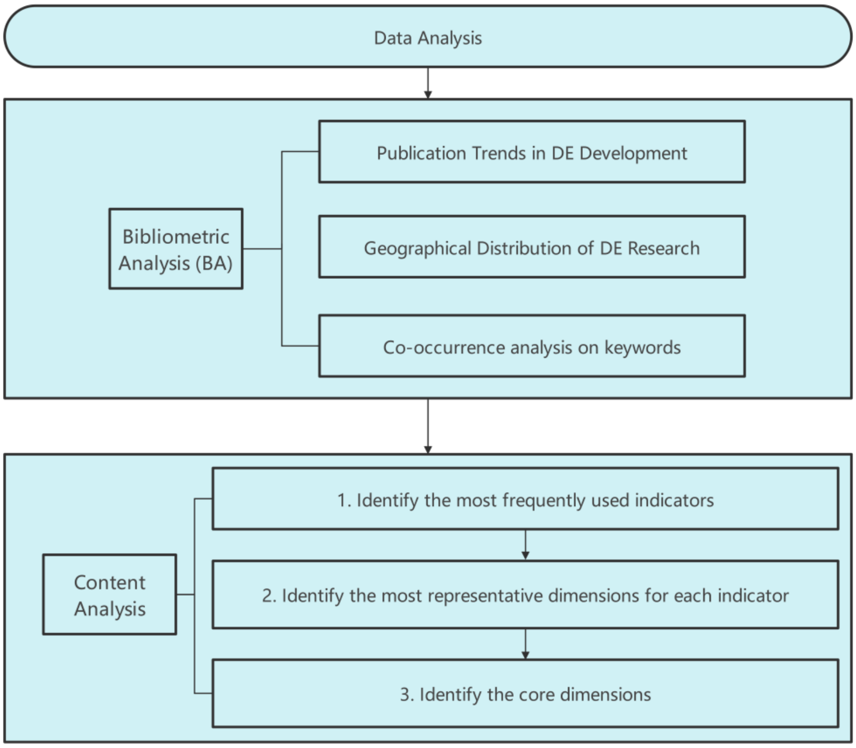


Figure 2. Flowchart of data analysis

Bibliometric analysis (BA)

By analysing publication data from the 161 identified studies, the patterns, trends and key research themes in the DE were identified. VOSviewer is utilised to create keyword co-occurrence networks. This software's ability to generate large, easily comprehensible maps makes it a suitable tool for this analysis (Donthu *et al.*, 2021; Kumar & Yadav, 2024).

Content analysis (CA)

This study conducted a CA of 161 selected articles to identify commonly used dimensions and indicators for measuring the DE. Initially, all dimensions and indicators mentioned in the papers were recorded and organised. Subsequently, a verbatim analysis was performed on these dimensions and indicators to assess their consistency and frequency of occurrence (Mengistu & Panizzolo, 2021). Indicators deemed fundamentally similar were combined and calculated together. The entire process was conducted in MS Excel and divided into the following three steps:

Step 1: Identify Most Frequently Used Indicators. To identify the most frequently used indicators, a threshold of 15 appearances was selected. Indicators exceeding this threshold will be identified and categorised into respective dimensions, based on their descriptions and classifications as outlined in the existing literature.

Step 2: Identify the Most Representative Dimension for Each Indicator. Indicators identified in Step 1 were categorised into dimensions based on existing literature. The most representative dimension for each indicator was determined by counting the frequency of its occurrence across different dimensions.

Step 3: Identify the Core Dimensions. Representative dimensions determined in Step 2 were ranked by frequency of occurrence to identify the core dimensions for assessing the DE. Figure 2 summarises the process involved in BA and CA.

BA Results

Publication trends in DE development

An analysis of 161 articles revealed distinct publication patterns over time. As shown in Figure 3, these trends can be visualised in three distinct stages, highlighting the evolution of the field.

The initial stage (2010–2016) saw a noticeable absence of research in DE development, indicating limited scholarly attention towards the development of the DE during this period. Moving into the second phase (2017–2020), the volume of literature revealed fluctuation with minimal overall growth. However, by 2020, a gradual increase in publications became evident, signalling a growing interest in DE research. The third phase (2021–2023) experienced a pronounced surge in published articles, reflecting an exponential growth trajectory in scholarly interest. Although comprehensive statistics for 2024 are not available yet, it is anticipated that scholarly literature on the DE will continue to expand throughout the year.

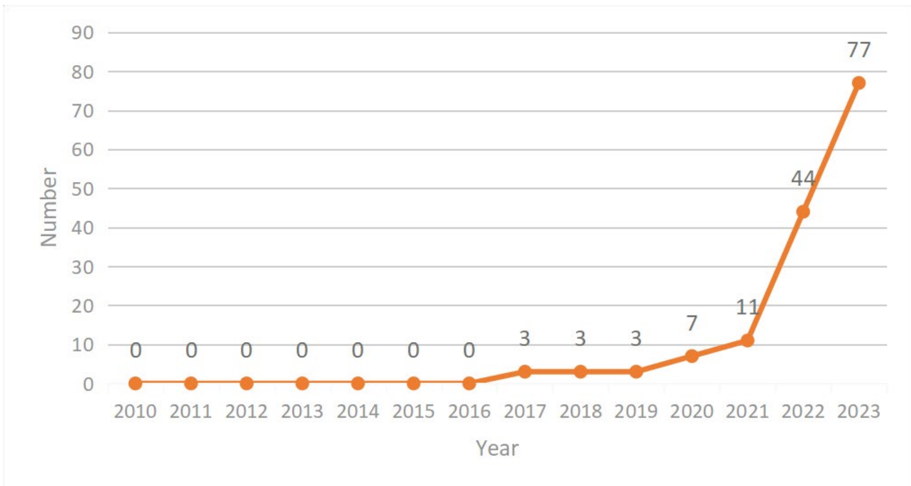


Figure 3. Publication of research papers related to the DE from 2010 to 2023

https://doi.org/10.18080/jtde.v12n3.988

Geographical distribution of DE research

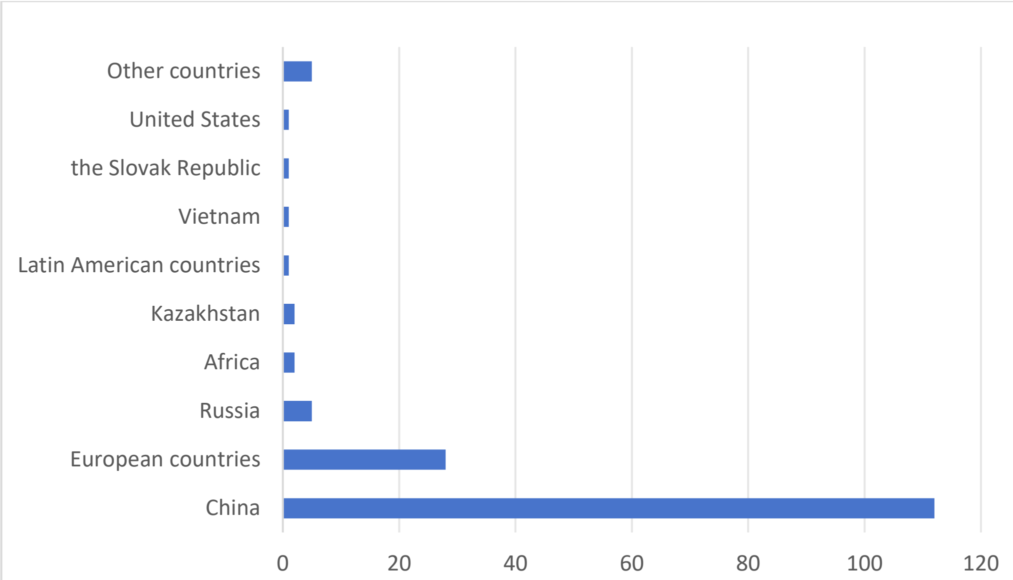


Figure 4. Geographical distribution of DE research

Figure 4 illustrates the geographic distribution of the 161 publications. The location of the study came from 10 countries or regions. China was the dominant contributor, representing approximately 70% of the total (112 articles). European countries collectively accounted for 28 articles, while Russia contributed five. In recent years, China has experienced rapid development in its DE. The *Digital China Development Report (2023),* published in June 2024, indicates that the scale of China's DE surpassed 55 trillion yuan in 2023. Additionally, the added value of the core industries within the DE constituted approximately 10% of the country's GDP (China National Data Administration, 2024). The increase in publications on China's DE is closely linked to the Chinese Government's policy initiatives. Between 2012 and 2021, the number of policies related to the DE rose from 19 to 335 (Qin *et al.*, 2023), driving a significant rise in research on the DE in China.

Co-occurrence analysis on keywords

Keywords serve as pivotal components of research, encapsulating the essence and scope of the study undertaken by the author. In this study, keyword co-occurrence analysis was conducted utilising VOSviewer software. Keywords that appeared more than five times in VOSviewer analysis were captured, yielding a final set of 52 keywords from the initial 903. This approach allowed for a systematic analysis of the interconnectedness and recurring themes within the relevant literature. The five most frequently occurring keywords were: (a) "digital economy" (142 times); (b) "China" (63 times); (c) "economic development" (34 times); (d) "economic and social effects" (25 times); and (e) "panel data" (25 times).

Figure 5 illustrates the keyword trends over time, offering valuable insights into the evolving landscape of research within the DE domain. Prior to 2021, there were relatively few publications pertaining to the DE. Hence, this section commences the analysis from the year 2021. Each node in the network represents a keyword, with its colour indicating the time of its appearance. Nodes progressed from blue colour on the left to red on the right. Red nodes represent keywords or topics of recent interest, while blue nodes signify those that were popular in the early stages of the DE development.

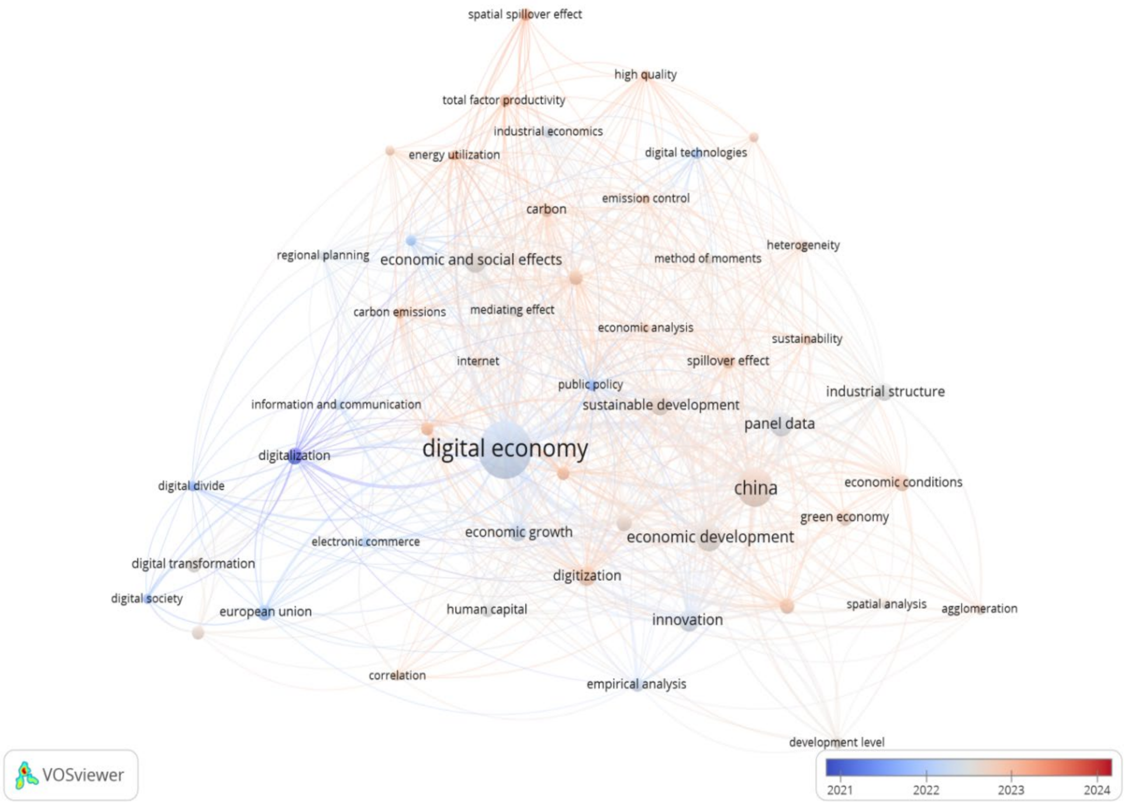


Figure 5. Keywords analysis with time information

Prior to 2022, DE research centred on foundational concepts and elements such as e-commerce, digitisation and the digital divide. This early phase reflected a fundamental understanding of the emerging economic landscape. Specifically, research related to the EU emphasised the pioneering role of developed nations in shaping early discussion of the DE.

After 2022, research in DE witnessed a surge, particularly from developing countries like China. This coincided with a clear shift in focus, moving beyond foundational concepts to examine the broader economic and social impacts. Themes like industry structure, economic development and carbon emissions emerged as key areas of investigation, reflecting a growing recognition of digitisation's multifaceted consequences. This shift signifies a maturing field, with researchers increasingly venturing into the complex interplay between the DE, society, industry and the environment.

https://doi.org/10.18080/jtde.v12n3.988

Figure 5 reflects the dynamism of research in the DE domain. It reveals a shift from foundational explorations of digital technologies to a more in-depth investigation of their far-reaching economic and societal consequences.

CA Results

Among the 161 articles, 148 papers focused on assessing the development of the DE. Among them, 33 utilised established indicators to measure DE development. The top three indicators, based on their usage frequency, are: DESI (used 23 times), NRI (5 times), and the Global Innovation Index (GII) (4 times). The remaining 115 articles opted to construct their own indicator systems to assess DE development.

A total of 48 indicators appeared at least three times across the 115 articles. Table 1 lists the 14 most frequently cited indicators (occurring at least 15 times) and their corresponding dimensions. Scholars exhibited diverse views on categorising these indicators into dimensions. To prioritise dimensions with broad applicability, this study selected the most frequently assigned dimension for each indicator, marked with an asterisk (*) in Table 1, as the representative dimension.

Table 1. Top 14 indicators and corresponding dimensions

	Indicator	Number of occurrences	Dimension (number of appearances)	Percentage
1	Mobile phone penetration rate	87	*Digital infrastructure (25)	28.7
			Internet development (22)	25.3
			Others (19)	21.8
			No classification (21)	24.1
2	Proportion of employees in information transmission, computer services and software	70	*Internet development (15)	21.4
			Digital industry development (11)	15.7
			Digital industrialisation (9)	12.9
			Digital talents (3)	4.3
			Others (14)	20
			No classification (14)	20
3	Internet penetration rate	70	*Digital infrastructure (24)	34.3
			Internet development (22)	31.4
			Others (17)	24.3
			No classification (7)	10
4	Telecommunications business volume per capita	63	*Internet development (17)	27
			Digital industrialisation (12)	19
			Digital industry (7)	11.1
			Information development (6)	9.5
			Others (14)	22.2
			No classification (7)	11.1

https://doi.org/10.18080/jtde.v12n3.988

	Indicator	Number of occurrences	Dimension (number of appearances)	Percentage
5	E-commerce sales	33	*Industrial digitisation (11)	33.3
			Digital transaction (7)	21.2
			Digital application (5)	15.2
			Others (9)	27.3
			No classification (1)	3
6	Number of domain names	28	*Digital infrastructure (16)	57.1
			Internet development (3)	10.7
			Others (8)	28.6
			No classification (1)	3.6
7	Number of Internet broadband access ports	28	*Digital infrastructure (15)	53.6
			Others (5)	17.9
			No classification (8)	28.6
8	Length of fibre optic cable lines	25	*Digital infrastructure (19)	76
			Others (4)	16
			No classification (2)	8
9	Software business revenue	24	*Digital industrialisation (11)	45.8
			Others (9)	37.5
			No classification (4)	16.7
10	Proportion of enterprises with e-commerce transactions	23	*Industrial digitisation (6)	26.1
			Digital transaction (4)	17.4
			Digital application (3)	13
			Others (7)	30.4
			No classification (3)	13
11	Number of patent applications	18	*Digital innovation (11)	61.1
			Others (7)	38.9
12	Number of Internet Broadband Access Users	16	*Digital infrastructure (4)	25
			Internet development (3)	18.8
			Others (5)	31.2
			No classification (4)	25%
13	Number of websites per 100 companies	16	*Industrial digitisation (5)	31.2
			Digital transaction (4)	25
			Others (7)	43.8
14	Number of computers per 100 people	14	*Industrial digitisation (2)	14.3
			Digital application (2)	14.3
			Others (9)	64.3
			No classification (1)	7.1

Each indicator corresponds to a representative dimension. In summary, it was found that the 14 indicators are distributed across five representative dimensions, namely digital infrastructure, industrial digitisation, Internet development, digital industrialisation and digital innovation. These five representative dimensions were ranked based on their frequency of occurrence to identify the core dimensions. Digital infrastructure emerged as the most prominent, appearing six times across the 14 dimensions, industrial digitisation (four times) and Internet development (two times). Digital industrialisation and digital innovation each appeared once.

https://doi.org/10.18080/jtde.v12n3.988

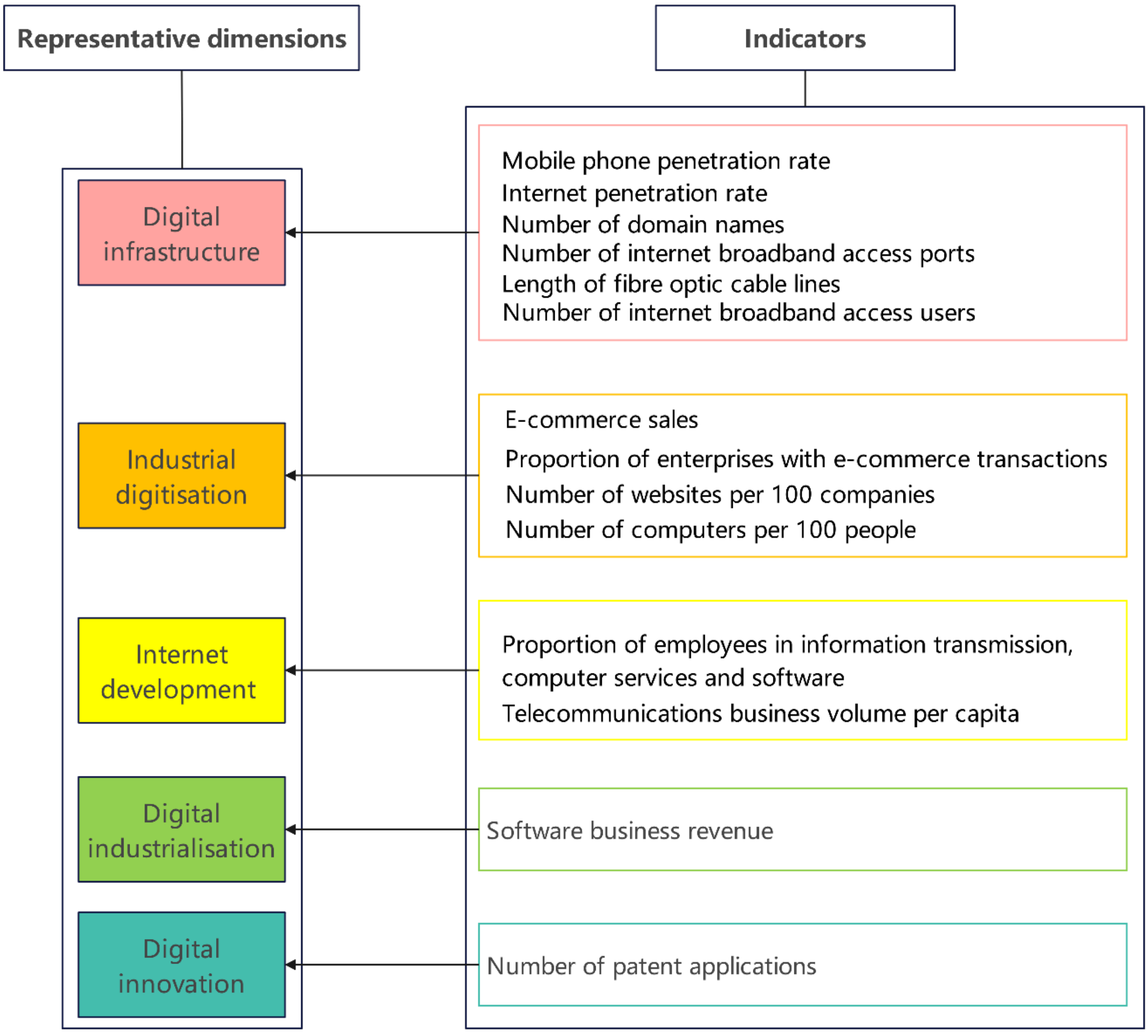


Figure 6. Selected representative dimensions and indicators

As shown in Figure 6, six indicators were classified under the representative dimension "digital infrastructure", including: mobile phone penetration rate; Internet penetration rate; number of domain names; number of Internet broadband access ports; length of fibre optic cable lines; and number of Internet broadband access users.

Four indicators were classified under the representative dimension "industrial digitisation", namely: e-commerce sales; proportion of enterprises with e-commerce transactions; number of websites per 100 companies; and the number of computers per 100 people.

The representative dimension "internet development" included two indicators: proportion of employees in information transmission, computer services and software, and telecommunications business volume per capita. "Digital industrialisation" and "digital innovation" each included two indicators: software business revenue and number of patent applications, respectively.

Discussion

The analysis of the 14 most frequently cited indicators revealed five dimensions closely associated with the development of the DE, namely digital infrastructure, industrial digitisation, Internet development, digital industrialisation and digital innovation. However, significant overlap exists among Internet development, digital infrastructure and digital industrialisation. For instance, mobile phone and Internet penetration rates were often categorised under both Internet development and digital infrastructure. To address these overlaps and build upon the work of Liu *et al.* (2024), Zhang *et al.* (2022a), and Liang *et al.* (2024), this study proposes a framework centred on four core dimensions: digital infrastructure, digital industrialisation, industrial digitisation and digital innovation.

Then the previously extracted indicators in 115 articles were categorised into four core dimensions based on their definitions and alignment with existing literature. To establish a comprehensive and balanced indicator system, the selection process considers both the frequency and recency of indicator appearance.

Digital infrastructure

Digital infrastructure is a critical prerequisite for the DE, facilitating the transmission and application of digital information and technology (Xing *et al.*, 2023). It unveils availability and quality of physical and virtual infrastructure supporting digital activities. Digital infrastructure serves as a novel form of an infrastructure system facilitating the digitalisation of both the economy and society (Rong, 2022). Numerous scholars, in their construction of DE development indicator systems, prioritise DE infrastructure as a crucial dimension (Liang *et al.*, 2024). This dimension is often integrated with others to comprehensively assess DE development.

Table 2. Key indicators of digital infrastructure

Indicators	**Number of occurrences(n=286)**	**Percentage**
Mobile phone penetration rate	87	30.4
Internet penetration rate	70	24.5
Number of Internet broadband access ports	28	9.8
Number of domain names	28	9.8
Length of fibre optic cable lines	25	8.7
Others (Number of active IP4 addresses etc.)	48	16.8
Total	**286**	**100**

Note: Number of occurrences means the number of times the indicator was used in 115 articles.

The preceding analysis highlights the critical role digital infrastructure plays as the foundation of DE. Table 2 delineates the top five indicators pertaining to digital infrastructure. Evidently, the leading indicator is the “mobile phone penetration rate”, constituting approximately 30.4%

of the overall usage frequency. Following closely is the "internet penetration rate", commanding approximately 24.5% of utilisation. Subsequently, "number of internet broadband access ports" assumes significance, representing approximately 17.5% of the total usage.

Digital industrialisation

Digital industrialisation constitutes a critical element of the DE. It refers to the use of digital technology to transform factors of production into digital products, and, as the size of such products grows, various new digital industries are formed (Liu & Ji, 2023). The absence of ICT technologies would have precluded the existence of these industries, such as telecommunications business and software business. Digital industrialisation underscores the industrialisation of digital technologies. Given the perpetual innovation and evolution of technology, from a narrow perspective digital industrialisation specifically focuses on the ICT sector. The broad definition encompasses a wider array of categories, incorporating domains such as digital content and transactions.

Table 3. Key indicators of digital industrialisation

Indicators	Number of occurrences (n=236)	Percentage
Telecommunications business volume per capita	63	26.7
E-commerce sales	33	14
Software business revenue	24	10.2
E-commerce purchases	10	4.2
Others (Information service revenue etc.)	106	44.9
Total	**236**	**100**

Note: Number of occurrences means the number of times the indicator was used in 115 articles.

According to the definition of digital industrialisation, industries that emerged following the advent of IT are categorised into this dimension (Table 3), such as telecommunication business, software business, express business, e-commerce, and information services. After counting these indicators, this study selects the top four indicators concerning digital industrialisation. The indicator with the highest frequency count is "telecommunications business volume per capita," accounting for 26.7% of the total occurrences. Following is "e-commerce sales," comprise approximately 14% of the dataset. "Software business revenue" is the third most frequent indicator, constituting 10.2% of the total occurrences. E-commerce purchases refer to the total value of goods and services acquired through online orders. Both e-commerce purchases and e-commerce sales are crucial components of e-commerce statistics and together constitute the turnover of e-commerce (Wang *et al.*, 2024).

Indicators related to telecommunications, e-commerce and software development feature prominently due to the longevity and maturity of ICT industries which have accumulated substantial data over their developmental trajectories. Conversely, promising digital industries, such as artificial intelligence, lack comparable historical data or maturity levels, thus presenting challenges in gathering extensive datasets for analysis.

Industrial digitisation

Industrial digitisation, on the other hand, refers to the utilisation of digital technology and data resources to improve output and efficiency in traditional industries, culminating in the integration of digital technology with the real economy. The concept of industrial digitisation shares commonalities with "digitalised economy". The latter refers to the economic activities that were established prior to the widespread adoption of ICT but are progressively integrating digitised data into their organisational processes (Ganichev & Koshovets, 2021).

Based on the above definition, this study categorises all indicators related to the use of IT to enhance the output and efficiency of traditional industries into the dimension of industrial digitisation.

Table 4 showcases the top four indicators in terms of frequency of use among the indicators related to industrial digitisation. The most frequently utilised indicator is the "proportion of enterprises with e-commerce transactions," representing 21.5% of the total usage. The "number of enterprise-owned websites" emerges as the second indicator, constituting 20.6% of the dataset. Ranked third in frequency is the "number of websites per 100 companies," accounting for 14.9% of the indicators utilised. As technology and the DE advance, an increasing number of enterprises are embracing this trend by harnessing Internet platforms for e-commerce, accelerating innovation in products and services, and precipitating organisational change, ultimately facilitating the digital transformation of these enterprises (Ren *et al.*, 2022).

Table 4. Key indicators of industrial digitisation

Indicators	Number of occurrences (n=107)	Percentage
Proportion of enterprises with e-commerce transactions	23	21.5
Number of enterprise-owned websites	22	20.6
Number of websites per 100 companies	16	15
The number of computers per 100 people	14	13.1
Others (Number of enterprises etc.)	32	29.8
Total	**107**	**100**

Note: Number of occurrences means the number of times the indicator was used in 115 articles.

Digital innovation

Table 5. Key indicators of digital innovation

	Indicators	Number of occurrences (n=130)	Percentage
Innovation output	Number of patent applications	17	13.1
	Technology market turnover	5	3.8
Innovation input	Proportion of employees in information transmission computer services and software	70	53.8
	R&D expenditure/GDP	8	6.2
Others	Expenditure in technological and science research, etc.	30	23.1
Total		**130**	**100**

Note: Number of occurrences means the number of times the indicator was used in 115 articles.

Innovation is a critical driver of enterprise competitiveness in the digital age (Budiarto & Nordin, 2024). Digital innovation, recognised as a critical dimension for assessing the DE, has gained scholarly attention in recent years (Liu *et al.*, 2024; Su *et al.*, 2023). Digital innovation can be categorised into two main components (Table 5): innovation input and innovation output (Zhang *et al.*, 2022b). Based on the study by the World Intellectual Property Organization (WIPO) in 2020, innovation input refers to the resources and efforts invested in the innovation process, such as human resources and financial investment. Innovation output is the tangible and intangible results derived from these inputs, including new products and services, and intellectual property (WIPO, 2020).

From an innovation output perspective, patent applications (13.1%) and technology market turnover (3.8%) are commonly utilised indicators, highlighting the market advantages derived from new technologies and products. Technology market turnover reflects the intermediate output of technological innovation (Hua *et al.*, 2022) and serves as an indicator of digital innovation (Su *et al.*, 2023). In terms of innovation investment, the prevalent indicators include the proportion of employees in information transmission, computer services, and software industries (53.8%), along with R&D expenditure as a percentage of GDP (6.2%), underscoring the significance of human capital and R&D funding.

Compared to previous research, this research emphasises the critical role of digital innovation in the DE. The OECD's Digital Economy Outlook 2020 (OECD, 2020) supports this by proposing patent applications as a key indicator. While the EU's DESI index doesn't explicitly measure digital innovation, it indirectly addresses this through human capital and technology adoption metrics (European Commission, 2022).

Building on previous research, this study proposes a framework for assessing DE development, outlined in Table 6. The framework comprises four dimensions and 17 indicators.

https://doi.org/10.18080/jtde.v12n3.988

Table 6. A framework to assess the development of DE

Core Dimension	Description	Indicators (institutions including or measuring the indicators)
Digital infrastructure	Availability and quality of physical and virtual infrastructure supporting digital activities	Mobile phone penetration rate (OECD; EU; ITU) Internet penetration rate (OECD; EU; ITU) Number of Internet broadband access ports Number of domain names (OECD) Length of fibre optic cable lines (ITU)
Digital industrialisation	New digital industries resulting from the emergence and development of digital technologies	Telecommunications business volume per capita E-commerce sales (EU; ITU) Software business revenue E-commerce purchases (EU; ITU)
Industrial digitisation	Utilisation of digital technology and data resources to improve output and efficiency in traditional industries	Proportion of enterprises with e-commerce transactions (OECD) Number of enterprise-owned websites (OECD) Number of websites per 100 companies The number of computers per 100 people (OECD)
Digital innovation	Innovation input and output	Proportion of employees in information transmission computer services and software (EU) R&D expenditure / GDP (OECD) Number of patent applications (OECD) Technology market turnover

Note(s):
(OECD): Organisation for Economic Co-operation and Development (2020). OECD Digital Economy Outlook 2020
(EU): European Union (European Commission, 2022). The Digital Economy and Society Index 2022 (DESI)
(ITU): International Telecommunication Union (2024). Measuring digital development: Facts and Figures 2023

Conclusion

Although the concept of the DE was explored early on, its measurement has comparatively lagged behind. With the introduction of an index by international organisations, a notable acceleration in academic literature on DE measurement has begun to emerge. Scholars have approached DE measurement by either adopting existing measurement indicators or modifying measurement frameworks based on their own interpretations, thereby contributing to the evolving discourse on this topic.

This paper provided a comprehensive overview into the research on the measurement of the DE, traced the evolving trends in the distribution of publications concerning DE measurement over time to offer a fundamental understanding of its evolution within the academic landscape. Through the VOSviewer analysis on principal keywords and topics, this study identified key research areas and focal points within the domain of the DE, as well as their chronological evolution. Further, the study statistically analysed indicators and primary dimensions employed in existing literature. The study employed SLR, BA and CA to conclude a framework system for assessing the development of DE across four dimensions: digital infrastructure, digital industrialisation, industrial digitisation and digital innovation. By integrating diverse metrics and indicators, this system provides a comprehensive, reasonable and generally applicable evaluation framework, facilitating informed decision-making and policy formulation in the digital age.

The framework proposed in this study offers several significant contributions to the assessment of DE development:

Data availability for quantitative analysis: While some scholars have proposed comprehensive frameworks for measuring DE, these frameworks often face limitations due to the unavailability of data, hindering subsequent quantitative analysis (Shostak *et al.*, 2023). This study addressed the data availability challenge by extracting indicators from 115 DE studies, and systematically analysing these indicators as well as identifying potential data sources for each indicator. This process established a solid foundation for subsequent quantitative analysis of DE development.

Comprehensive indicator system with multi-dimensional coverage: Previous studies typically assessed the DE through two or three dimensions, such as informatisation, Internet and digital transactions (Zhang *et al.*, 2023), or digital infrastructure, digital industrialisation and industrial digitisation (Li & Zhao, 2023). This study developed a more comprehensive indicator system by incorporating four core dimensions: digital infrastructure, digital industrialisation, industrial digitisation and digital innovation. The study evaluated and integrated the indicators and dimensions used in prior research, thus ensuring a broader and more inclusive measurement framework.

Scientific and rational basis: This study employed SLR to integrate and summarise previous research findings, identifying indicators that are both widely used and empirically validated. This approach minimises the impact of individual subjective judgements, thereby providing strong empirical support for the scientific and rational foundations of the proposed indicator system.

The scientific and comprehensive nature of the proposed framework serves as a valuable tool for policymakers, researchers and practitioners seeking to understand and measure DE progress. By providing a data-rich, comprehensive, and scientifically grounded approach, this framework supports informed decision-making and drives sustainable digital economic growth.

The main limitation of the study is the insufficient incorporation of emerging indicators identified in recent literature. Consequently, the proposed framework may not fully capture the dynamic nature of DE development. Future research should address this gap by

https://doi.org/10.18080/jtde.v12n3.988

integrating these indicators into the measurement framework to enhance its alignment with the dynamic DE landscape and improve the accuracy of development assessments.

References

Aktoprak, A., & Hursen, C. (2022). A bibliometric and content analysis of critical thinking in primary education. *Thinking Skills and Creativity*, *44*, 101029. https://doi.org/10.1016/j.tsc.2022.101029

Anthopoulos, L. G., & Tzimos, D. N. (2021). Carpooling platforms as smart city projects: A bibliometric analysis and systematic literature review. *Sustainability*, *13*(19), 10680. https://doi.org/10.3390/su131910680

Atkinson, R. D. (January 19, 2021). A U.S. Grand Strategy for the Global Digital Economy. Information Technology & Innovation Foundation. https://itif.org/publications/2021/01/19/us-grand-strategy-global-digital-economy/

Baquee, A., Rahaman, M. S., & Sevukan, D. (2023). A bibliometric review of academic social networking sites (ASNSs) in scholarly communication: a scientific mapping based on Scopus database. *International Journal of Information Science and Management (IJISM)*, *21*(3), 286–309. https://doi.org/10.22034/ijism.2023.1977938.0

Bruno, G., Diglio, A., Piccolo, C., & Pipicelli, E. (2023). A reduced Composite Indicator for Digital Divide measurement at the regional level: An application to the Digital Economy and Society Index (DESI). *Technological Forecasting and Social Change*, *190*, 122461. https://doi.org/10.1016/j.techfore.2023.122461

Budiarto, D. S., & Nordin, N. (2024). Technology transformation, innovation, and digital economy development in developing countries: A systematic literature review. *Journal of Telecommunications and the Digital Economy*, *12*(1), 148–171. https://doi.org/10.18080/jtde.v12n1.802

Bukht, R., & Heeks, R. (2017). Defining, conceptualising and measuring the digital economy. Development Informatics Working Paper 68. http://dx.doi.org/10.2139/ssrn.3431732

CAICT. (2017). *White Paper on China's Digital Economy Development*. China Academy of Information and Communications Technology. http://www.caict.ac.cn/english/research/whitepapers/202104/t20210429_375940.html

CAICT. (2023). China Digital Economy Development Report. China Academy of Information and Communications Technology. http://www.caict.ac.cn/kxyj/qwfb/bps/202207/P020220729609949023295.pdf

Cheng, Y., Zhang, Y., Wang, J., & Jiang, J. (2023). The impact of the urban digital economy on China's carbon intensity: Spatial spillover and mediating effect. *Resources, Conservation and Recycling*, *189*, 106762. https://doi.org/10.1016/j.resconrec.2022.106762

China National Data Administration. (2024). *Digital China Development Report (2023)*. https://www.digitalchina.gov.cn/2024/xwzx/szkx/202406/P020240630600725771219.pdf

Davidson, A. (2015, November 9). The Commerce Department's Digital Economy Agenda. *Worth Magazine*. https://techonomy.com/the-commerce-departments-digital-economy-agenda/

Department for Digital, Culture, Media & Sport. (October 4, 2022). Ministerial foreword and executive summary. https://www.gov.uk/government/publications/uks-digital-strategy/uk-digital-strategy.

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, *133*, 285–296. https://doi.org/10.1016/j.jbusres.2021.04.070

European Commission. (2022). Digital Economy and Society Index (DESI) 2022. https://digital-strategy.ec.europa.eu/en/library/digital-economy-and-society-index-desi-2022

G20 Research Group. (2016). G20 Digital Economy Development and Cooperation Initiative. https://www.mofa.go.jp/files/000185874.pdf

Ganichev, N. A., & Koshovets, O. B. (2021). Forcing the digital economy: How will the structure of digital markets change as a result of the COVID-19 pandemic. *Studies on Russian Economic Development*, *32*, 11–22. https://doi.org/10.1134/S1075700721010056

Hanna, N. K. (2020). Assessing the digital economy: Aims, frameworks, pilots, results, and lessons. *Journal of Innovation and Entrepreneurship*, *9*(1), 1–16. https://doi.org/10.1186/s13731-020-00129-1

Henry, D. K., Buckley, P., Gill, G., Cooke, S., Dumagan, J., & Pastore, D. (1999). *The Emerging Digital Economy II*. United States Department of Commerce Washington, DC. https://www.commerce.gov/sites/default/files/migrated/reports/ede2report_0.pdf#:~:text=This%20emerging%20digital%20economy%20regularly%20surprises%20those%20who%20study%20it

Hua, Q., Liu, G., Sun, D., & Zhu, L. (2022). Spatial effects of technology market development on energy efficiency: Heterogeneity analysis based on the characteristics of technology market. *Technological Forecasting and Social Change*, *185*, 122008. https://doi.org/10.1016/j.techfore.2022.122008

Irkinovich, N. R. (2022). The digital economy today. *Academicia Globe: Inderscience Research*, *3*(10), 198–203. https://agir.academiascience.org/index.php/agir/article/view/938

International Telecommunication Union. (2024). Measuring digital development: Facts and Figures 2023. https://www.itu.int/en/ITU-D/Statistics/Pages/facts/default.aspx?os=ios&ref=app#:~:text=Steady%20but%20uneven%20progre%20ss%20in%20global%20Internet%20connectivity%20highlights

Jiang, X. (2020). Digital economy in the post-pandemic era. *Journal of Chinese Economic and Business Studies*, *18*(4), 333–339. https://doi.org/10.1080/14765284.2020.1855066

Khan, M. N., & Abdullah (2021). Determining mobile payment adoption: A systematic literature search and bibliometric analysis. *Cogent Business & Management*, *8*(1), 1893245. https://doi.org/10.1080/23311975.2021.1893245

Knickrehm, M., Berthon, B., & Daugherty, P. (2016). Digital disruption: The growth multiplier. *Accenture Strategy*, *2016*(1), 1–11. http://www.metalonia.com/w/documents/Accenture-Strategy-Digital-Disruption-Growth-Multiplier.pdf

Kumar, N. K., & Yadav, A. S. (2024). A systematic literature review and bibliometric analysis on mobile payments. *Vision*, *28*(3), 287–302. https://doi.org/10.1177/09722629221104190

Lange, S., Pohl, J., & Santarius, T. (2020). Digitalization and energy consumption. Does ICT reduce energy demand? *Ecological Economics*, *176*, 106760. https://doi.org/10.1016/j.ecolecon.2020.106760

Li, Q., & Zhao, S. (2023). The impact of digital economy development on industrial restructuring: Evidence from China. *Sustainability, 15*(14), 10847. https://doi.org/10.3390/su151410847

Liang, B., He, G., & Wang, Y. (2024). The digital economy, market integration and environmental gains. *Global Finance Journal*, *60*, 100956. https://doi.org/10.1016/j.gfj.2024.100956

Liu, H., & Ji, R. (2023). The mechanism and effect of the promotion of industrial structure upgrading by digital economy. *Science & Technology Progress and Policy*, *01*, 61–70 (In Chinese). https://doi.org/10.6049/kjjbydc.2022060328

Liu, J., Fang, Y., Ma, Y., & Chi, Y. (2024). Digital economy, industrial agglomeration, and green innovation efficiency: empirical analysis based on Chinese data. *Journal of Applied Economics*, *27*(1), 2289723. https://doi.org/10.1080/15140326.2023.2289723

Ma, Q., Mentel, G., Zhao, X., Salahodjaev, R., & Kuldasheva, Z. (2022). Natural resources tax volatility and economic performance: Evaluating the role of digital economy. *Resources Policy*, *75*, 102510. https://doi.org/10.1016/j.resourpol.2021.102510

Mengistu, A. T., & Panizzolo, R. (2021). Indicators and framework for measuring industrial sustainability in Italian footwear small and medium enterprises. *Sustainability*, *13*(10), 5472. https://doi.org/10.3390/su13105472

Millar, J. N., & Grant, H. (2019). Valuing the digital economy of New Zealand. *Asia-Pacific Sustainable Development Journal*, *1*, 1–19. https://dx.doi.org/10.18356/2a2fdd33-en

Moulton, B. R. (2000). GDP and the digital economy: Keeping up with the changes. In *Understanding the Digital Economy: Data, Tools, and Research* (pp. 34–48). https://dx.doi.org/10.7551/mitpress/6986.003.0004

Nagy, S., & Somosi, M. V. (2022). The relationship between social innovation and digital economy and society. *Regional Statistics*, *12*(2), 1–27. https://dx.doi.org/10.15196/RS120202

OECD. (2020). OECD Digital Economy Outlook 2020. OECD Publishing. https://www.oecd.org/en/publications/2020/11/oecd-digital-economy-outlook-2020_3f7b7e58.html

Pan, W., Xie, T., Wang, Z., & Ma, L. (2022). Digital economy: An innovation driver for total factor productivity. *Journal of Business Research*, *139*, 303–311. https://doi.org/10.1016/j.jbusres.2021.09.061

Portulans Institute. (2023). Network Readiness Index 2023. https://networkreadinessindex.org/nri-2023-edition-press-release/

Pulsiri, N., & Vatananan-Thesenvitz, R. (2018). Improving Systematic Literature Review with Automation and Bibliometrics. 2018 Portland International Conference on Management of Engineering and Technology (PICMET), 1–8. IEEE Explore. https://doi.org/10.23919/PICMET.2018.8481746

Qin, M., Mirza, N., Su, C. W., & Umar, M. (2023). Exploring bubbles in the digital economy: The case of China. *Global Finance Journal*, *57*, 100871. https://doi.org/10.1016/j.gfj.2023.100871

Rehman, N. U., & Nunziante, G. (2023). The effect of the digital economy on total factor productivity in European regions. *Telecommunications Policy*, *47*(10), 102650. https://doi.org/10.1016/j.telpol.2023.102650

Ren, Y., Zhang, X., & Chen, H. (2022). The impact of new energy enterprises' digital transformation on their total factor productivity: Empirical evidence from China. *Sustainability*, *14*(21), 13928. https://doi.org/10.3390/su142113928

Rong, K. (2022). Research agenda for the digital economy. *Journal of Digital Economy*, *1*(1), 20–31. https://doi.org/10.1016/j.jdec.2022.08.004

Shahbaz, M., Wang, J., Dong, K., & Zhao, J. (2022). The impact of digital economy on energy transition across the globe: The mediating role of government governance. *Renewable & Sustainable Energy Reviews*, *166*, 112620. https://doi.org/10.1016/j.rser.2022.112620

Shostak, L., Goi, V., Timchenko, O., Yastrubetska, L., & Derhaliuk, M. (2023). The impact of digital transformation on the economy: technological innovation and efficiency. *Revista de Gestão e Secretariado (Management and Administrative Professional Review)*. https://doi.org/10.7769/gesec.v14i10.2883

Singh, V. K., Singh, P., Karmakar, M., Leta, J., & Mayr, P. (2021). The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis. *Scientometrics*, *126*, 5113–5142. https://doi.org/10.1007/s11192-021-03948-5

State Council of the People's Republic of China, The. (2023, February 28). China unveils plan to promote digital development. https://english.www.gov.cn/policies/latestreleases/202302/28/content_WS63fd33a8c6d0a757729e752c.html.

Su, J., Dong, C., Su, K., & He, L. (2023). Research on the construction of digital economy index system based on K-means-SA algorithm. *Sage Open*, *13*(4). https://doi.org/10.1177/21582440231216359

Tapscott, D., & Caston, A. (1994). Paradigm shift: The new promise of information technology. 62–66. McGraw-Hill, Inc. Professional Book Group, 11 West 19th Street, New York, NY, United States.

Tripathi, M., & Dungarwal, M. (2020). Digital India: Role in development. *International Journal of Home Science*, *6*(2), 388–392. https://www.homesciencejournal.com/archives/2020/vol6issue2/PartG/6-2-72-511.pdf

United States Department of Commerce. (2016, June 22). Enabling Growth and Innovation in the Digital Economy. https://www.ntia.gov/report/2016/enabling-growth-and-innovation-digital-economy.

Wang, J., Wang, B., Dong, K., & Dong, X. (2022). How does the digital economy improve high-quality energy development? The case of China. *Technological Forecasting and Social Change*, *184*, 121960. https://doi.org/10.1016/j.techfore.2022.121960

Wang, Y., Song, G., & Wang, T. (2024). Evaluation of the level of digital economy development and its spatial-temporal evolution analysis in the Yellow River Basin. *Yellow River*, *46*(05), 11–17. [Chinese]. https://doi.org/10.3969/j.issn.1000-1379.2024.05.002

Williams, L. D. (2021). Concepts of digital economy and industry 4.0 in intelligent and information systems. *International Journal of Intelligent Networks*, *2*, 122–129. https://doi.org/10.1016/j.ijin.2021.09.002

World Intellectual Property Organization (WIPO). (2020). The Global Innovation Index 2020 – Appendix 1. Retrieved from https://www.wipo.int/edocs/pubdocs/en/wipo_pub_gii_2020-appendix1.pdf

Xing, Z., Huang, J., & Wang, J. (2023). Unleashing the potential: exploring the nexus between low-carbon digital economy and regional economic-social development in China. *Journal of Cleaner Production*, *413*, 137552. https://doi.org/10.1016/j.jclepro.2023.137552

Yang, S., & He, J. (2022). [Retracted] Analysis of digital economy development based on AHP-Entropy Weight Method. *Journal of Sensors*, *2022*, 1–8. https://doi.org/10.1155/2022/7642682

Yang, Y. (2023). The impact of the digital economy on young people's consumption in the context of the new coronary pneumonia epidemic, *Economic Research-Ekonomska Istraživanja*, *36*(3), 2212743. https://doi.org/10.1080/1331677X.2023.2212743

Zhang, W., Zhao, S., Wan, X., & Yao, Y. (2021). Study on the effect of digital economy on high-quality economic development in China. *PloS one*, *16*(9), e0257365. https://doi.org/10.1371/journal.pone.0257365

Zhang, H., Dong J., & Wang L. (2022a). High-quality development of China's digital economy: Connotation, current situation and countermeasures. *Journal of Humanity*, *318*(10), 75–86. [Chinese] https://doi.org/10.15895/j.cnki.rwzz.2022.10.009

Zhang, J., Lyu, Y., Li, Y., & Geng, Y. (2022b). Digital economy: An innovation driving factor for low-carbon development. *Environmental Impact Assessment Review*, *96*, 106821. https://doi.org/10.1016/j.eiar.2022.106821

Zhang, W., Zhang, S., Bo, L., Haque, M., & Liu, E. (2023). Does China's regional digital economy promote the development of a green economy? *Sustainability*, *15*(2), 1564. https://doi.org/10.3390/su15021564

Zhang, K., Cao, B., Guo, Z., Li, R., & Li, L. (2024). Research on the impact of government attention on the digital economy of Chinese provinces. *Innovation and Green Development*, *3*(2), 100118. https://doi.org/10.1016/j.igd.2023.100118

An Early Management System for Subscriber Equipment Installation

Simon Moorhead
Telecommunications Manager

Abstract: The *Journal* revisits an historic paper from 1970 detailing the introduction of a computer-based Subscriber Installation Management Control System.

Keywords: History of Australian Telecommunications, Subscriber Installation Management Control System, SIMS

Introduction

The historic paper (Keighley & Higgins, 1970) details the introduction of a computer-based Subscriber Installation Management Control System (SIMS) in Adelaide by the Australian Post Office (APO, now Telstra) in 1970. The paper was chosen as an early example of a computer management system in Australia and complements the special-issue section on "Perspectives on Machine Learning" in this issue of the *Journal*.

Subscriber equipment installation and rearrangement was the mainstay of services provided by the APO in 1970. It involved two different staff groups: lineman installers under the local Lines organisation; and technician installers under divisional management control. The technician installer activities present supervision problems, given the installers are in small groups dispersed over wide areas and have a wide variety of equipment and building types to deal with.

The SIMS system was designed to improve the scheduling of field staff, balance the work loads of groups and depots in the division, detect non-productive effort and time, and assess work content in significant categories for management purposes. The paper states "the proposed system allows labour resources to be accurately matched to work loadings and provides meaningful information to supervisors and engineers to pinpoint areas and conditions that may warrant improvement".

https://doi.org/10.18080/jtde.v12n3.1013

The working time needed to be divided into direct and indirect work categories. The indirect work was further divided into lost time and overheads. An Activity Definition Manual was developed and the acceptable times for activities were workshopped and agreed with representatives of the technicians and management. These agreed times were centrally controlled for accuracy reasons and could be applied with confidence by field personnel. The installers were also issued with a pocket-sized handbook to facilitate the correct coding prior to reporting information to the depot.

The SIMS system had two fundamental segments, namely the work scheduling phase and the information system. The depots filled out CE60 computer source documents detailing the work achieved for the period and these were forwarded to the divisional Costing Section. Duplicate CE60s were forwarded to a computer service bureau for transcription into computer punch cards.

Some of our readers may not be familiar with computer punch cards but they were the input interface to mainframe computers at the time. Computer service bureaux ran programs on their computers on behalf of customers (often overnight or over several days), then couriered the output results by way of large printouts (the striped fanfold pages, which were standard at that time) back to their customers. The SIMS system provided four types of output reports: performance, costing, survey and control.

After the 12-month trial period, the SIMS system could boast of making available information to operational personnel for the previous period, within three working days of the end of that period. This seems extraordinary, given today's "real time" resource management systems. Divisional and supervisory management were also provided with sufficient information to enable them to assess performance and initiate corrective action, where necessary. Work scheduling and the reallocation of resources were made easier, as well as the identification of inefficient procedures.

The SIMS system was designed and implemented within the South Australian division of the PostMaster General's Department (PMG), to be used by all APO subscriber equipment installers in the PMG at that time. The aims of achieving better planning and control of work were realised, as well as the most efficient utilisation of the available resources.

References

Keighley, R., & Higgins, P. (1970). The Subscriber Installation Management Control System. *Telecommunication Journal of Aust*ralia, *20*(2), 135–143.

https://doi.org/10.18080/jtde.v12n3.1013

The Historic Paper

June, 1970 THE TELECOMMUNICATION JOURNAL OF AUSTRALIA 135

THE SUBSCRIBERS INSTALLATION MANAGEMENT CONTROL SYSTEM

R. KEIGHLEY, A.R.M.T.C. and P. HIGGINS, B.E., Dip.Eng.Mngt.***

INTRODUCTION.

In the Australian Post Office the work of installing or rearranging telephone equipment in subscribers' premises is carried out by two different staff groups. Linemen-installers undertake the less complex installations, whilst technician-installers undertake the more complex work. The latter group work entirely on internal equipment and operate from subscribers' installation depots with a line of control through senior and supervisory technicians to an engineer with overall responsibility for the Subscribers' Installation function. The former group work on both internal and external plant, and operate from what is usually called a Line Depot with a line of control through Line Foremen and Line Inspectors usually to an Engineer with overall responsibility for External Plant installation and maintenance (District Works).

Technician-installer activities, which range over a wide variety of equipments and building types, present the usual supervision problems which arise when several small work groups are dispersed over wide areas at locations which are changing frequently. (Ref. 1).

The system described in this article was designed to test means of improving the supervision and management of technician-installer activities.

SYSTEM OUTLINE.

The objectives of the system are to supplement and strengthen the supervision of the depot supervising and senior technicians and to provide them, and engineering management, with an information and control system having the following features:

(i) Scheduled allocation of work to field staff.
(ii) Work load balancing between working groups and depots within a division.
(iii) Detection of non-productive effort and time.
(iv) Work content assessment in significant categories for management purposes.

The proposed system allows labour resources to be accurately and continually matched to work loadings and provides meaningful information to supervisors and engineers to pinpoint areas and conditions which may warrant improvement.

WORK MEASUREMENT.

In order to initiate the required work measurement, the total working time was broken down into direct and indirect work. The indirect work was further divided into lost times and overheads whilst the direct work was divided into clearly defined elements called activities. The total time to perform a task is also dependent on such variables as environment, class of service (provide, remove or disconnect) and type of service (internal or external). In the case of environment, a sample survey led to the description of five types of building situations, from single-storey houses or villa residences to multi-storey houses, offices and factories.

The Activity Definition Manual is an important document in the establishment of acceptable times for activities employed in the system. It must be centrally controlled to ensure that the integrity of the derived information is maintained and that the full system remains meaningful to management. The manual is divided into Commonwealth standard activities and activities of local (State) significance only.

The activity definitions have been made available at field level in two different formats. The full manual is used to estimate major and minor works, to schedule work to the installers, and to calculate work load figures for staff balancing. A condensed pocket-size handbook is issued to all installers for quick reference to identify tasks and determine the correct coding prior to reporting information to the depot.

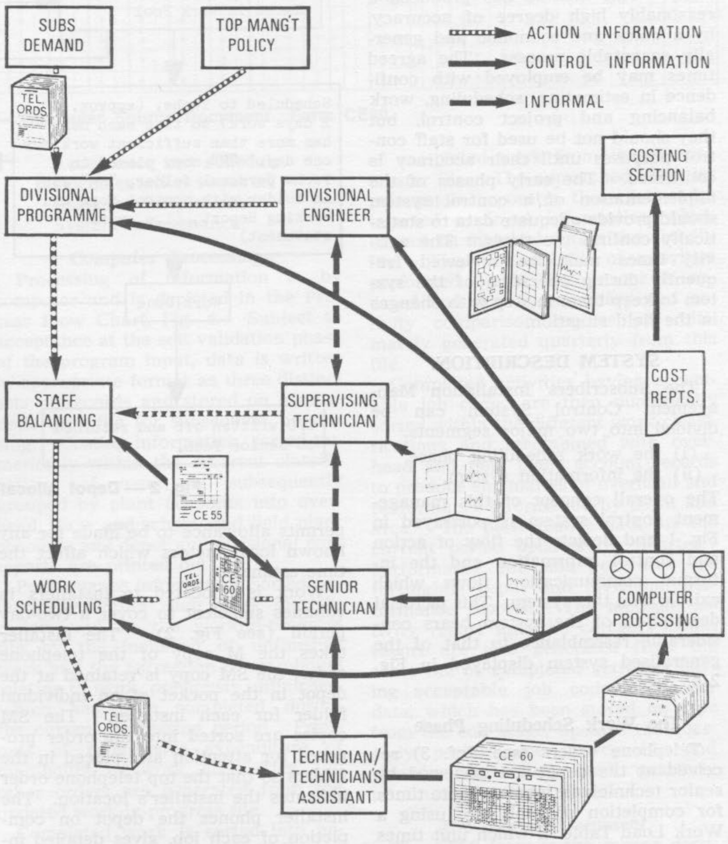


Fig. 1 — Subscribers Installation Management Information and Control System (SIMS).

* Mr. Keighley is Engineer Class 3, Industrial Engineering and Training, Headquarters. See Vol. 20, No. 1, p. 82.
**Mr Higgins is Engineer Class 2, Industrial Engineering and Training, Headquarters. See Vol. 20, No. 1, p. 82.

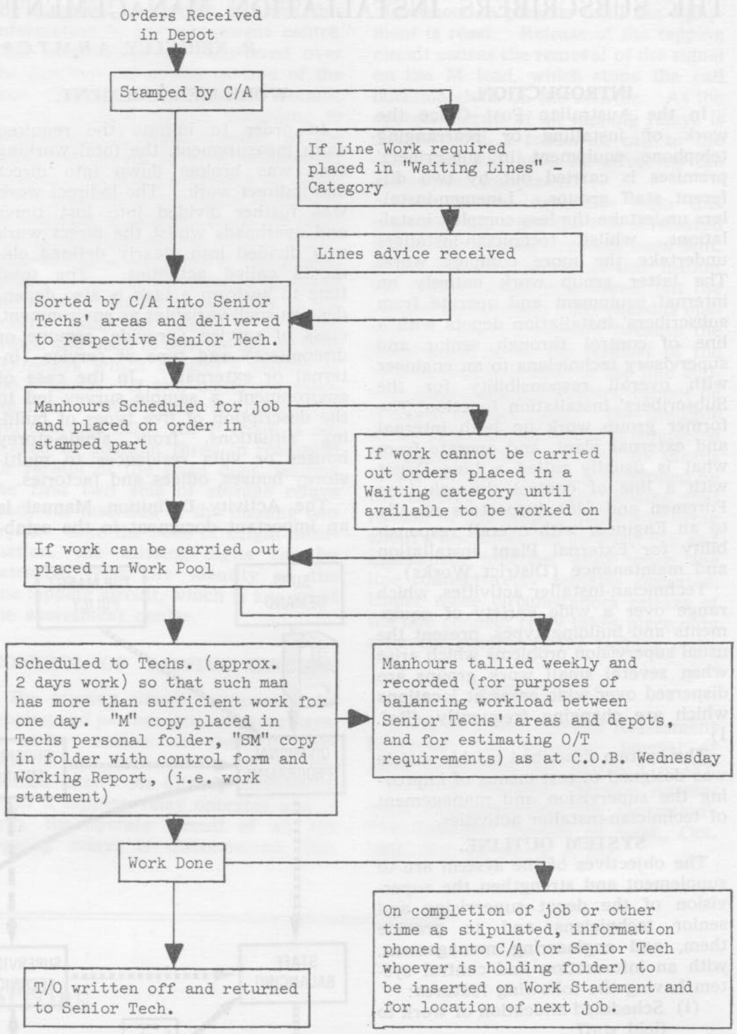


Fig. 2 — Depot Allocation of Telephone Orders.

The activity manuals are used initially to establish acceptable performance times for each defined activity and these times will ultimately be used for job allocation and management. The activity times have been determined by the 'agreed times conference' technique. Representative technicians, senior technicians, supervising technicians and engineers who are familiar with the activities being discussed and who are capable of assessing the work content, form the conference, which is chaired by an impartial officer who is familiar both with the activities and the requirements of work measurement, generally an Industrial Engineer.

The procedures of the conference require that all representatives consider each activity in turn, confirm the definition and then agree to an estimate of the normal time taken by a qualified technician, operating under normal conditions, to perform the activity. This method has produced a reasonably high degree of accuracy, by a quick and economic and generally acceptable process. The agreed times may be employed with confidence in estimating, scheduling, work balancing and project control, but they should not be used for staff control purposes until their accuracy is confirmed. The early phases of the implementation of a control system should provide adequate data to statistically confirm the times. The activity times must be reviewed frequently during operation of the system to keep them in step with changes in the field situation.

SYSTEM DESCRIPTION.

The Subscribers' Installation Management Control System can be divided into two major segments.

(i) the work scheduling phase,
(ii) the information system.

The overall concept of this management control system is portrayed in Fig. 1 and depicts the flow of action and control information and the informal communication flows which exist within this system. This pictorial description of the system bears considerable resemblance to that of the generalised system displayed in Fig. 2 of Ref. 2.

The Work Scheduling Phase

Telephone orders (see Ref. 3) received at the depot are reviewed by senior technicians, who estimate times for completion of each job, using a Work Load Table in which unit times are derived from the basic activity times. This facilitates scheduling and permits allowance to be made for any known local factors which affect the time estimate.

Work is allocated to installers in volumes sufficient to cover a two-day period (see Fig. 2). The installer takes the M copy of the telephone order; the SM copy is retained at the depot in the pocket of an individual folder for each installer. The SM copies are sorted into the order proposed for attention and placed in the folders so that the top telephone order indicates the installer's location. The installer phones the depot on completion of each job, gives detailed information about the job and nominates his next location; these details are transcribed on to form CE60 (see Fig. 3), which is held in each installer's folder.

The CE60 form, in the initial phases of the system development, is completed in duplicate, the carbon copy (which replaces the working report WP1) being subsequently forwarded to Costing Section, as the work statement, at the end of each period. The SM copy of the order is marked 'completed' and placed beneath the incomplete orders in the folder so that the uppermost order, clearly visible in the pocket, is the one for the job in hand. M and SM copies of completed telephone orders are subsequently reunited for normal depot pro-

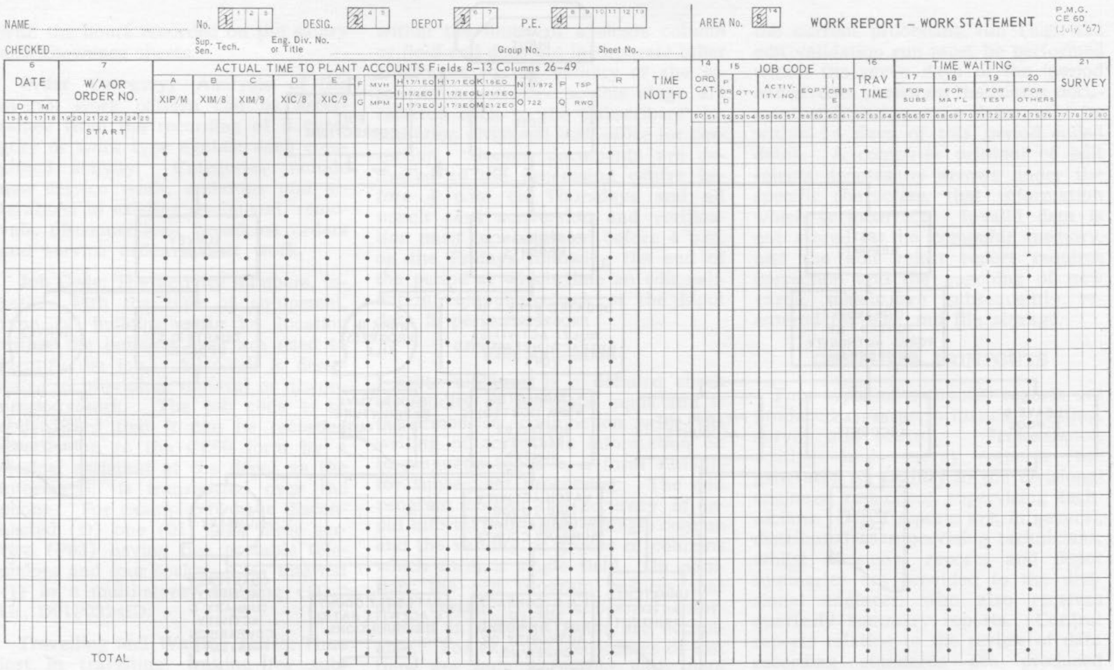


Fig. 3 — Computer Source Document. Form CE60.

cessing and despatching. A yellow slip is associated with the M copy and provides a record of information passed to the depot by the installer. Sample checks by the senior technicians ensure that installers are coding accurately and providing all the required data.

At the end of each week the CE60 forms are checked by the respective senior technicians and the two copies are forwarded to the divisional office, whence one copy is routed to the Costing Section, the other to the computer service bureau for transcription to computer punch cards.

The Information System.

The information system has been developed to meet the Subscribers' Installation Measurement System (SIMS) specification — Version 'A,' which details the requirements of data processing and includes the following information:—

- System description
- Source document and associated input forms
- Punching instructions
- Computer punch card verification schedule
- System analysis/flow process diagram
- Processing requirements
- File description
- Computer printout formats
- Acceptance test schedule
- Documentation required
- Basis for calculations
- System maintenance

Computer Processing.

Processing of information is by computer and is depicted in the Process Flow Chart, Fig. 4. Subject to acceptance at the edit validation phase of the program input, data is written in appropriate format as three distinct sets of records and stored on a magnetic tape file for subsequent processing. Costing information, sorted numerically within the different classifications of work, is subsequently grouped by plant accounts into overhead, leave and school, and field plant accounts and depot and divisional cost reports are printed out.

Performance information, sorted numerically within each work classification, updates the work-in-progress file by the inclusion of unfinished work and the extraction of man-hours relating to work, previously unfinished, which has been completed during the current processing period. Printouts of progressive expenditure of man-hours to work authorities and other orders are generated from the updated work-in-progress file.

Completed work is sorted into activity number sequence and the activity time is matched to each record of completed work from the activity master file. The unmatched activities printout incorporates those job codes inserted as input information which cannot be matched within the activity file. The summary file of activities performed within the quarter is updated and activity summary and activity comparison reports are ultimately generated quarterly from this file.

Completed activities having acceptable job codes are then numerically sorted within the different work classifications and recombined with overhead and field plant account records to produce summaries of overhead and performance information by depot and area. The performance data for the current period updates the quarterly performance file and generates the fortnightly depot and divisional performance reports. The multiple activity report and the category count report are also generated from the same file of completed activities having acceptable job codes. Survey data, which has been stored on tape from the edit stage, updates the quarterly performance survey file and provides fortnightly and quarterly printouts of survey statistics.

The Source Document.

The CE60 form is the source document from which data is transcribed

Journal of Telecommunications and the Digital Economy, ISSN 2203-1693, Volume 12 Number 3 September 2024

https://doi.org/10.18080/jtde.v12n3.1013

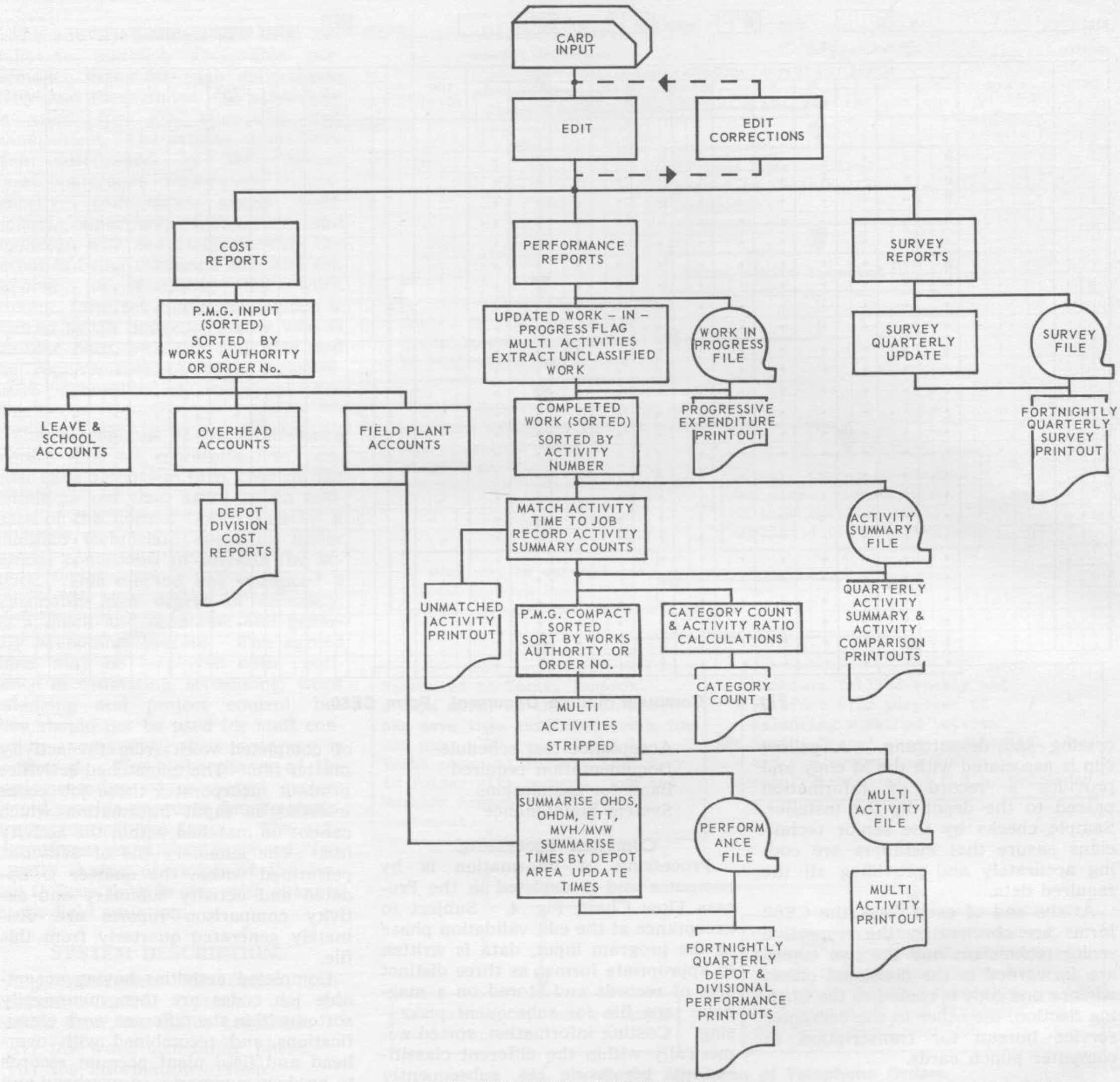


Fig. 4 — SIMS Process Flow Chart.

to computer punch cards as the computer input. Personal identification information is recorded in the header of the form. The body of the form makes provision for specific and detailed entries peculiar to each job performed.

Particular reference should be made to certain information fields of the source document.

Designation: Technicians-in-training are distinguished from other installation staff by the insertion of designations T1 to T5 for first to fifth year trainees. Other staff are designated T0. Designation factors are specified for each in accordance with costing procedures (T1 = 0; T2 = 0.4; T3 = 0.55; T4 = 0.70; T5 = 0.85 and T0 = 1.0), such that the effective work performed is assessed by multiplying actual time expended on the activity by the appropriate designation factor.

W/A Or Order No.: The telephone order or works authority number describes the activity on which work is being performed. Differentiation between the various works authorities and different types of telephone orders is possible by self-explanatory prefixes to the numbers. Supervisory overheads (OHDS) and miscellaneous overhead activities (OHDM), excess travelling time (ETT) and time expended on vehicles denoted by the vehicle number (e.g., BH 345) can also be recorded in this field as desired.

Actual Time to Plant Accounts: Times are recorded in hours to one decimal place in the appropriate plant account columns. Some columns make provision for alternate plant account entries and for the insertion of other plant accounts as desired. The recording of plant account times by columns permits their ready summation by the Costing Section. The daily hours recorded for each member of staff on the CE60 form must coincide

KEIGHLEY & HIGGINS — Subscribers Installation Management Control System with the hours recorded on the salary and allowance sheets.

Order Category: Analysis of the overall work pattern can be ascertained from the recording of a category or work type against each completed activity. Categories include new service work, additions and alterations to existing equipment, removals, disconnections, broadcast orders and service order docket work.

Job Code: The activity numbers, together with the type of equipment, type of building and the location within the building area are used to identify the job. All work is designated as providing, removal or disconnection work. The job code is so constructed that it can be recorded directly on to the CE60 form by a clerical assistant at the depot as the information is transmitted over the phone. For example, "I have disconnected (D) one (01) plan three service (P03) having two portable telephones and four sockets (24) internal (I) in a multi-storey building (D)," i.e., D01P0324ID.

Travelling and Waiting Times: Time lost in travelling, waiting for subscribers, waiting for material and waiting test are recorded on the source document. A further column 'waiting others' provides management with the facility to study any particular aspect of lost time desired.

Survey: Columns have been made available to permit the special study of any desired item of interest by the insertion in these columns of data information which can be processed, under direction to the service bureau, to provide the desired printout. The original survey performed has studied the times expended in testing subscribers' services using the exchange test desk by comparison with line test robots.

PROCESSING OF SOURCE DOCUMENTS

Source documents are processed and the various computer printouts are generated fortnightly and quarterly and dispatched to the divisional office for distribution to supervising and senior technicians at each Subscribers' Installation depot. Source documents are initially converted to computer punch cards at the Service Bureau and are subsequently machine verified before input to the computer. A verification schedule for conversion to punch cards has been drawn up which refers to information omitted or outside specified limits. The verification phase is restricted to checking within the limits of a single column or field and does not incorporate other checks as to the validity of the information. Errors at this stage are referred back to the Subscribers' Installation divisional staff office for correction. Source documents are delivered to the processing centre before 4 p.m. on Thursdays and all punch card conversion and verification must be completed before 4 p.m. on the Fridays following the end of the period at which time an edit validation computer check of the input data is performed.

Edit Validation

Edit validation is a detailed check of validity of the data input and is performed in conjunction with the writing of acceptable information on to magnetic tapes in a form suitable for subsequent processing. The edit program checks the accuracy of input information within a single column and the validity of groups of columns which form a single field. In addition, the edit program ensures that only certain combinations of information in different fields are acceptable. For example, recordings of lost time are only permitted with field plant accounts; leave and training school entries relate only to certain specified plant accounts, as do motor vehicle and overhead costs; works authority and telephone order identification numbers must meet specific constraints, and only certain combinations are acceptable in the formation of the job code. Time checks are incorporated into the edit validation run to ensure that only source documents for the previous period are included in the current processing run (e.g., the edit validation run must be performed within the seven days of the 'period ended' date shown on the source document and performance dates must be within 14 days of this period ended date). A computer printout of edit rejects denotes by arrows under the specific characters, that information which is incorrect. Invalid data is not stored in the computer memory, and the edit reject report requires correction and the punching of new cards, which are subsequently re-entered for edit and file storage.

COMPUTER PRINTOUTS

Four types of computer reports are produced — performance, costing, survey and control. Performance, costing and survey reports are generated fortnightly to the divisional engineer and depot supervising technicians. Depot reports are, in general, distributed to supervising technicians, whilst divisional reports and depot summaries are provided to the divisional engineer, who also receives quarterly summary reports. Control reports relating to the review of data recording standards are produced fortnightly and other control reports providing for the review of activity time standards are generated quarterly. The distribution of printouts is depicted in Fig. 5.

Costing Report.

Depot and divisional summaries of costing reports are generated fortnightly. Reports are divided into three sections — field plant accounts, overhead plant accounts and leave

REPORT		DIVISIONAL ENGINEER	SUPERVISING TECHNICIAN	SENIOR TECHNICIAN	COSTING SECTION	HEADQUARTERS
FORTNIGHTLY REPORTS	EDIT VALIDATION REPORT	*				
	UNMATCHED ACTIVITY REPORT	*				*
	COSTING REPORT	S	*		*	*
	PROGRESSIVE EXPENDITURE OF MANHOURS REPORT	*	*	*	*	*
	PERFORMANCE REPORT	S	*	*		*
	WORK TYPE REPORT	S	*			*
	SURVEY REPORT	S	*			*
QUARTERLY REPORTS	ACTIVITY SUMMARY REPORT	*				*
	ACTIVITY COMPARISON REPORT	*				*
	PERFORMANCE REPORT	S	*	*		*

S = SUMMARY

Fig. 5 — Distribution of Computer Print-outs.

https://doi.org/10.18080/jtde.v12n3.1013

and school accounts — and each is further subdivided to the actual plant accounts. Hours (adjusted for trainee contribution) are shown, together with cost expenditure to each plant account. The duplicate copy of the CE60 form provides information to Costing Section which was previously recorded on the WP1 forms, from which plant account expenditure can be processed by the Costing Section. Ultimately the costing report may be the only working document provided to the Costing Section and the duplicate CE60 form will not be needed. This by-product of the SIMS system could lead to considerable staff savings within the Costing area.

Progressive Expenditure of Man Hours to Works Authorities

Man-hours expended against works authorities are recorded progressively and the printout summary depicts man hours expended to date against each plant account (adjusted for trainee contribution) on current work. Progressive expenditure reports are provided each fortnight to depot supervising technicians and highlight the senior technician area(s) in which the work is being performed.

Unmatched Activity Report.

The activity codes and their associated activity times are stored in the computer's activity master file and activity codes recorded as input data are checked against this file. Only certain combinations of information which form the activity code are acceptable, and although each segment of the job code may in itself be correct, combinations which do not form acceptable job codes are incorporated in this Unmatched Activity Report.

This reject report occurs during the computer processing run and complete rejection of the data at this stage would result in the loss of considerable amounts of otherwise acceptable data, including actual manipulative times, essential to the reconciliation of costing information. Accordingly the source data should not be rejected and, in practice, the unacceptable job code is treated as having zero activity time for performance calculations and all times expended are incorporated in the costing reports.

The Unmatched Activity Report identifies the job and the depot from which the source document originated. Investigations may show a need for further training of the staff concerned.

A downward trend in total hours recorded against unmatched activities would indicate a steady improvement in the operation of the system by field staff.

Maintenance Update Report.

Amendments to the master file are punched on cards for input to the computer. Whenever the content of the master file is changed or the activity times within the master file are amended, a maintenance update report is generated, indicating the acceptable amendments performed to the activity master file, together with edit listing of errors relating to unacceptable input data introduced in the update operation.

Performance Report.

The divisional engineer is provided with a fortnightly summary of performance reports produced for each depot, which are subdivided within each depot report into senior technicians' areas. A typical depot performance report is depicted as Fig. 6.

Two measures of performance are generated, manipulative performance and overall performance. Manipulative performance — the comparison of standard activity times with the equi-

SUBSCRIBER INSTALLATION MEASUREMENT SYSTEM A D E L A I D E.

DEPOT PERFORMANCE REPORT CITY DEPOT PERIOD ENDING 26/03/69

PM60

AREA	ACTIVITY TIME	ACTUAL TIME	MANIP PERFORM XXXXXX %	TOTAL LOST TIME	TRAVEL TIME	SUBS	WAITING TIMES MATERIAL	TEST	OTHER
A	296,00	351.87	84%	25.50	20.60	1.10	0.80		3.00
B	566.50	610.95	93%	43.40	33.80	2.20	1.00		6.40
C	221.50	171.40	129%	27.70	21.50	1.00	1.50	1.80	1.90
D	203.25	206.20	99%	51.30	29.60	7.90	6.80		7.00
E	537.25	586.12	92%	84.90	44.20	24.20	4.60		11.90
F	156.00	115.36	135%	18.90	7.40	2.30	2.50		6.70
G	57.25	56.50	101%	5.00	3.50	1.50			
DEPOT TOTAL	2037.75	2098.40		256.70	160.60	40.20	17.20	1.80	36.90

OHDM TIME	599.20
OHDS TIME	792.00
ETT TIME	
UNCLASS ACTY TIME	334.29
VEHICLES TIME	
TOTAL	3823.89

DEPOT PERFORMANCE STATISTICS

OVERALL PERFORMANCE	53%
MANIPULATIVE PERFORMANCE	97%

Fig. 6 — Typical Computer Print-out.

valent actual times for completed jobs — is provided for each senior technician's area. Overall performance— which incorporates depot overheads — is produced for each depot. Actual times expended on defined activities, travelling and waiting times, are itemised for each senior technician's area, whereas supervisory and depot overheads, ETT and motor vehicle times, are summarised on a depot basis.

Category Report.

Completed activities are categorised by work types for each depot and division and the activity ratio provides a measure of the percentage distribution of time between the different categories. This report highlights the relative emphasis on the different work types for such diverse purposes as the measure of effectiveness, cost/benefit analysis and loading and grading. For example, the prime function and major work content of Subscribers' Installation divisions might be considered to be the installation of new services. Analysis in Adelaide indicates that the installation of new services accounts for 6 per cent., provision of additional equipment to an existing service, 54 per cent., and alterations to existing installations, 25 per cent. of the work performed by the technician-installer staff group.

Activity Summary Report

The quarterly activity summary shows the number of occurrences of each activity on each type of work performed. This report provides information which is useful for material ordering.

Activity Comparison Report

The quarterly report is designed as a control report for continuing review of activity times initially determined at the Agreed Times Conference. The average actual time and the frequency of occurrence of each activity are tabulated, together with the variation of actual time from the activity time. The variation index is specified as:—

$$\frac{\text{Activity Time — Actual Time}}{\text{Activity Time}}$$

PRESENTATION OF INFORMATION

Analysis of information generated by computer printouts is assisted by the graphical recording of such information and the review of trends to detect changes in the various measures of performance. Action to effect productivity improvement could result from the review of these trends and the instigation of analysis and subsequent action to arrest unsatisfactory trends. It must be emphasised at this point that the computer processing of information and subsequent transcription of printout information to graphical format will never in itself result in productivity improvement. Essential to any productivity improvement is the need to analyse the information in detail, compare actual results in relation to the previous trends and anticipated results and instigate corrective action, where necessary. Control is the essence of all productivity improvement and the necessity for continuing control must be emphasised at all levels of supervisory management. To this end a three-section compendium has been provided which, in addition to graphical recording, has provision for recording of statistical data relating to depot and divisional reports and a reference to the distribution of computer printouts. The statistical information is arranged in such a manner that divisional summary totals are readily portrayed down the right-hand column of each page, the widths of which have been staggered such that all divisional information is immediately visible. The equivalent information relating to each depot is obtained simply by raising the appropriate flap, where such information is shown in the left-hand portion of each page.

Senior technicians are encouraged to perform their own recording from the depot printouts and to analyse the information in order to instigate corrective action.

IMPLEMENTATION OF THE SYSTEM.

This system was implemented as a trial in the Subscribers' Installation Division (SID), Adelaide. Field procedural manuals were prepared and pre-implementation training courses established. The scheduling of work to installers and the subsequent recording of data on to forms CE60 was introduced from the second period of January, 1968. During the next five months, programming and testing were completed in accordance with the SIMS specification and processing of current data commenced from 4th July, 1968. During these interim months, training of recorders and installers proceeded, with information recorded on the CE60 form being continually checked, firstly by the respective depot senior technicians and subsequently by the specialist senior technician nominated to provide assistance in the development of the system. Errors and omissions detected were referred back to the originating source and further explanation and guidance provided as an integral part of the training of recorders and installers towards the generation of accurate and complete input information.

Acceptance testing of the computer program was satisfactorily completed during the first week in June and corrected information generated in the interim period, transcribed on to punch cards, was processed progressively at the rate of two to three computer runs per week during the remainder of the month, primarily to create the necessary historic data within the computer files. In addition, progressive processing of back data further checked the capability of the program to handle large volumes of data, to correctly process and update information and to generally accommodate the variety and combinations of information which might normally be experienced in practice.

System Stability.

The effectiveness of the system in increasing the productivity of Subscribers' Installation activities has been assessed only after confidence had been gained in the trial system and it had become stable. To achieve this confidence, four criteria were aimed at:

(i) Errors produced by human action should be reduced to a low level.
(ii) Installers should be conversant with the concept of correct coding of jobs.
(iii) The identification of activities should be competently performed by installers.
(iv) The initial standard activity times which form the basis of the control function should be reasonable and acceptable times for the defined activities.

Monitoring of these criteria has been provided by control reports designed into the information system. The control features inherent in the information system are as follows:

Input Edit: Rejects are listed during the edit validation run prior to normal processing and a statement of total error rejects, together with total cards read, is provided. The ratio of errors to cards read has decreased from over 10 per cent. during the early periods of the system to less than 1 per cent. since August, 1968.

Uncompleted Work: The volume of uncompleted telephone order work should remain fairly constant if staff levels and overtime are reasonably steady and this feature is used to ensure that completed work is being coded off as required in the design of the system.

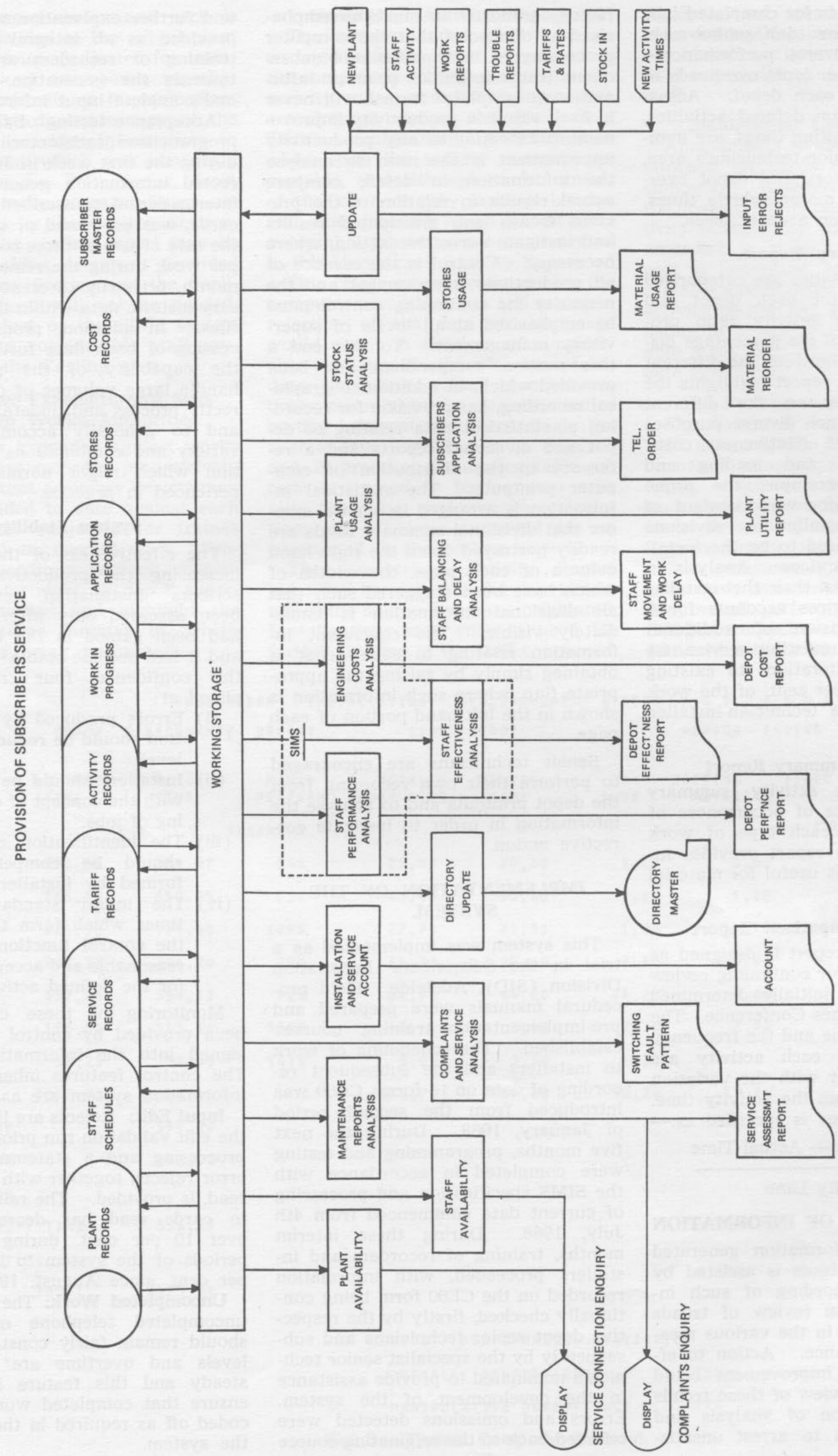


Fig. 7 — Possible Data Base for Comprehensive Management Information System, including SIMS. (See dotted area.)

During the early periods of the system many jobs were not being cleared by the reporting of the appropriate job code, which was indicated by a substantial buildup of uncompleted work. Job codes were subsequently inserted for these completed jobs and the volume of uncompleted work was thus reduced and has since remained fairly stable at about 5000 manhours per period.

Unmatched Activities: A number of codes are employed in describing a completed job, and the unmatched activity report depicts unacceptable codes. Reduction in the number of these indicates that the installers are translating the job identification into correct codes.

Activity Variation: The credibility of any control system rests on the accuracy of the standards employed in relation to the recorded data. Accordingly, the actual time for the activities are compared with the standard time and the variation of the average (expressed as a percentage) from this standard is listed. Variations greater than a predetermined limit indicate that the standard is at fault or that factors external to the system are affecting the activity times. This check is performed periodically. A variation index is employed to highlight the relativity of the actual to standard activity time, but because of the generally positive skewness of the statistical distribution of these recordings, the mode or median may be more appropriate.

RESULTS ACHIEVED.

Considerable benefits have been experienced by the management in Adelaide. Particular improvements achieved during the first 12 months of the trial were as follows:—

(i) Information relating to the previous period is available within three working days of the end of that period.

(ii) Divisional and supervisory management are provided with sufficient information to enable them to adequately assess performance and initiate corrective action when necessary.

(iii) Inefficient procedures which result in excessive time spent travelling, waiting for material, subscribers, test, etc., are highlighted.

(iv) Management is provided with costing details within three working days at the end of the period. Senior technicians also can more closely control expenditure on works authorities in progress.

(v) Supervising and Senior technicians become more management conscious.

(vi) Work scheduling can be performed more effectively, and staff locations can be readily ascertained from the depot by means of the work allocation procedures.

(vii) Facilities to effect work balancing and reallocation of staff reduce the build-up of backlogs of work at particular depots or areas.

CONCLUSION.

This article has described a complex management control system designed to suit the telephone subscribers' installation function. The system has taken just over three years from survey to audit of the trial situation, and present indications are that it will make a significant contribution to productivity improvement. The information generated by the system provides middle and top management with better opportunities to control their work, particularly in the following areas:—

(i) Budget control procedures may be implemented at sectional level.

(ii) The effect of the introduction of new equipment, methods or practices can be shown.

(iii) Staff loading and grading can be more accurately checked and updated.

(iv) Those areas of activity which offer the greatest prospect for productivity improvement can be identified.

(v) Material usage and therefore material ordering can be more accurate and stocks can be maintained at a lower level with safety.

The system has been designed with sufficient capacity to be used in all subscribers' installation units throughout the Commonwealth and with the information output integrated for the highest levels of management. Also, it was recognised that the information derived from the system is a small element in a larger, more comprehensive system for provision and service of the telephone subscriber (see Fig. 7).

The system aims at achieving better planning and control of work. The gain from the field staff point of view is reduction in frustrations at the work face caused by inadequate control using existing methods. Information made available to all levels allows all to participate in the continuing search for the most effective and efficient use of the available resources.

ACKNOWLEDGMENTS.

The development of this management control system has been the result of teamwork and the authors wish to acknowledge the contribution of their colleagues in the Industrial Engineering Unit, the Subscribers' Equipment and Telegraphs Section at Headquarters, the South Australian Administration and I.C.L.

REFERENCES

1. M. J. Power, 'Substation Installation Management and Practices in the Sydney Metropolitan Area'; Telecom. Journal of Aust., Feb. 1961, Vol. 12, No. 6, p. 451.
2. R. Keighley & P. Higgins, 'Management Control Systems;' Telecom. Journal of Aust., Feb. 1970, Vol. 20, No. 1, p. 74.
3. N. R. Cameron and K. N. Smith 'Telephone Orders by Teleprinter;' This issue, p. 144.

Further Reading

P. E. Rosove, 'Developing Computer-based Information Systems'; Wiley, New York, 1967.

B. Hodge & R. Hodgson, 'Management and the Computer in Information and Control Systems'; McGraw-Hill, New York, 1969.

R. O. Boyce, 'Integrated Managerial Controls'; Longmans, London, 1967.

R. H. Brandon, 'Management Standards for Data Processing'; Van Nostrand, New York, 1963.

B. Langefors, 'Theoretical Analysis of Information Systems'; Lund, Studentlitteratur, 1966.

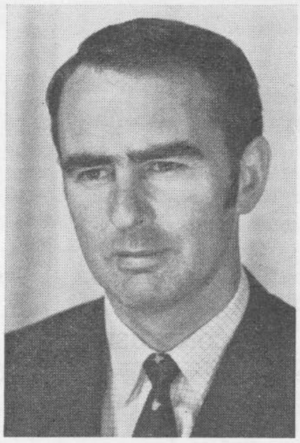
A. M. McDonough, 'Information Economics and Management Systems'; McGraw-Hill, New York, 1963.

https://doi.org/10.18080/jtde.v12n3.1013

From *Telecommunication Journal of Australia*, vol. 20, no. 1, p. 82:

OUR CONTRIBUTORS

P. J. HIGGINS



R. KEIGHLEY

P. J. HIGGINS, co-author of the article 'Management Control Systems', joined the Postmaster-General's Department as a Cadet Engineer in 1959. He graduated Bachelor of Electrical Engineering from Melbourne University in 1962, and completed a post-graduate Diploma of Engineering Management at Melbourne University in 1964. From 1963 he was Engineer Class 1 in Metropolitan Service Section, Victoria and was involved both with exchange maintenance activities and the development of the fault despatch centre and centralised testing at Ivanhoe. In May 1967, Mr Higgins was appointed Engineer Class 2 in the Industrial Engineering and Training Section, Headquarters where he has been closely associated with the development of management control systems for subscribers installation and exchange maintenance activities.

★

R. KEIGHLEY, co-author of the article 'Management Control Systems', joined the Postmaster-General's Department as a technician-in-training in 1950. After working as a technician on telephone exchange maintenance he qualified as senior technician and was appointed trainee engineer during 1957. He completed the Diploma of Communication Engineering in 1959 and was with the Victorian Metropolitan Installation Section until 1961. From 1961 to 1966 Mr Keighley was engaged at Headquarters, as an Engineer Class 2, in the design of telephone instruments and P.A.B.Xs. In February 1967 Mr Keighley was promoted to his present position as Engineer Class 3, Industrial Engineering, where he has been responsible for the design and development of management control systems as well as the co-ordination and development of computer applications within the Engineering Works Division.

★

https://doi.org/10.18080/jtde.v12n3.1013

Exploring Digital Proficiency Among Mothers of School-going Children in Kerala

A Comprehensive Analysis

V. S. Arya
School of Management Studies, Cochin University of Science and Technology, India

S. Sreejith
School of Legal Studies, Cochin University of Science and Technology

Pillai Rekha
Faculty of Business and Law, British University in Dubai, United Arab Emirates

Abstract: Early exposure to social and digital media in this digital age greatly influences children's lives, leading to an increasing number of parents and educators becoming concerned about the potential negative effects of excessive screen time. By investigating mothers' digital competencies as informal educators, this study fills a significant research gap. The paper emphasises women as mentors, highlighting their responsibility to supervise and guide their children's use of digital devices in an era of pervasive connectedness. Using a modified version of the DigCompEdu framework, the study focuses on two essential elements: empowering children; and fostering their understanding of digital competence areas. A thousand homemakers from two wards of the Kochi Corporation area in the Ernakulam district, Kerala, India are included in the study. Discriminant analysis is used to identify four important skill areas that demonstrate significant group differences – between Novice, Basic, Independent, and Proficient user groups. Results imply that tailoring digital literacy programs to mothers' varying proficiency levels can enhance their ability to navigate digital parenting effectively. Additionally, integrating sociodemographic factors into interventions can ensure relevance and efficacy in addressing digital competence disparities among mothers.

Keywords: Digital Technology Adoption, Digital Parenting, Digital Competence, Proficiency Level, Discriminant Analysis

https://doi.org/10.18080/jtde.v12n3.969

Introduction

In today's interconnected world, digital parenting is essential for guiding children's safe and responsible use of the Internet and digital devices (Gür & Türel, 2022). This responsibility begins as early as infancy, due to the widespread adoption of touchscreen devices among younger children (Mascheroni & Ólafsson, 2014; Chassiakos *et al.*,2016). Even toddlers as young as 2-4 years old are proficient in using tablets or smartphones to engage in games or watch videos, often encouraged by parents seeking to manage their behaviour in challenging social settings (Mascheroni & Ólafsson, 2014). As digital technology permeates various aspects of daily life, parents and children are increasingly exposed to media experiences, leading families to integrate digital technologies into activities, such as entertainment, education, and problem-solving. Consequently, navigating the digital landscape has become integral to modern family dynamics, highlighting the importance of fostering digital literacy and responsible digital citizenship from an early age.

Parents are the first mediators of their children's interactions with digital technologies. They are responsible for including their use in daily activities while promoting positive and secure interactions (Benedetto & Ingrassia, 2020). Digital parenting, including online safety education, screen time restrictions, tackling cyberbullying, and fostering healthy online interactions, are all tasks that parents can help with. Parents can assist children in navigating a wealth of information available online by teaching them media literacy and critical thinking techniques (Gür & Türel, 2022; Sela *et al.*, 2020; Vale *et al.*, 2018; Carvalho *et al.*, 2015). They can also create a safe environment for children to share their online experiences and concerns, providing guidance and supporting educational and creative uses of technology (Gür & Türel, 2022; Sela *et al.*, 2020; Vale *et al.*, 2018). This paper focuses on Kerala, a pioneering state in India known for pro-poor policies and social protection programmes for children and women in the most marginalised communities (Biswas & Narendranath, 2015). As of December 2020, India had 700 million active Internet users aged two years and above, which describes the country's dynamic digital ecosystem (Mohan *et al.*, 2020). The increase is notable in rural areas, emphasising the need to understand mothers' digital parenting abilities to supervise their children's usage of digital devices. However, children's usage of digital technology has dramatically increased in Kerala in recent years, posing both opportunities and concerns. Seventy-five per cent of children between the ages of 6 and 18 possessed access to smartphones, tablets, or computers, according to a 2021 survey by the Kerala State Commission for Protection of Child Rights. The way children learn and communicate has been revolutionised by this widespread access to technology but has also, in parallel, raised concerns regarding screen time and Internet safety (Nagy *et al.*, 2023; Chandrasekaran & Purushothaman, 2022). Nonetheless, in Kerala, mothers have been crucial in helping their children manage the digital world (Kallarakal & Gonsalvez, 2021; Nair *et al.*, 2007). With 68% of mothers indicating that they frequently communicate with their children about Internet safety, many mothers are actively involved in monitoring their children's online activity (Kallarakal & Gonsalvez, 2021; Nair *et al.*, 2007; Chacko, 2003). Additionally, they are encouraging educational apps and utilising technology to enhance the education of their children.

Despite having a strong digital vision, Kerala has experienced a sharp rise in cyber-attacks and security breaches over the past several years, with a significant portion of its population being victims of cybercrime (Kallarakal & Gonsalvez, 2021). The state has fallen into Internet addiction among children as a result of the widespread availability of Android phones and the Internet. An expert study by the Department of Child and Family Welfare in 2019 has noted that lack of family bonding, lack of extended family support, time mismanagement of working parents, children's incessant digital technology use, and single parenting are on the rise, all of which has led to reduced care and attention towards school-going children who need close monitoring and guidance (Kallarakal & Gonsalvez, 2021; Cingel & Kcrmar, 2013; Chacko, 2003). At this point, we argue that mothers, in particular, have a greater impact on children's social and intellectual behaviour. Kerala boasts a high rate of digital penetration and a tech-savvy population, making it pertinent to understand how parents, especially mothers, navigate the digital landscape concerning their children's digital technology use (Kallarakal & Gonsalvez, 2021; Nair *et al.*, 2007). However, the digital competence of mothers in Kerala, a state known for its high literacy rates and technological advancement, can vary depending on various socio-demographic factors. By identifying the different digital competency levels of school mothers in managing children's digital technology use, the research intends to provide insights into the most successful approaches for digital parenting in the Kerala context, ultimately contributing to the development of tailored guidelines and interventions for parents, schools, and policymakers to enhance the digital wellbeing of school children in the state. Therefore, this paper aims to achieve twofold objectives: firstly, to identify the digital competence of mothers of school-going children in Ernakulam District, Kerala; and classify maternal digital competence based on their digital parenting capabilities.

The paper contributes to the digital parenting domain in several ways. This novel study on mothers' digital competency in Kerala's Ernakulam District marks a pioneering exploration into digital parenting in a region renowned for technological innovation and digital prowess. Considering Kerala's socio-cultural nuances, these findings lay the groundwork for tailored guidelines and interventions catering to the local community. Beyond academic contributions, the study holds practical significance, potentially influencing parenting behaviours by offering strategies to regulate children's digital lifestyles. Additionally, the study may influence policy by enabling evidence-based recommendations and support initiatives for parents within the complex digital landscape. This research, the first of its kind in Kerala, proves to be a significant turning point in understanding and tackling the difficulties associated with digital parenting in this technologically advanced area.

The paper is structured as follows: the first section is the introduction section, followed by a review of the literature on digital competence and the framework for digital competence; the third section includes the research methodology, research context, survey design, and analysis design; the fourth section elaborates on the survey results; the fifth section offers a discussion of the results and implications of the study; and the final section presents the conclusion of the research and potential directions for future research.

Theoretical Framework and Literature Review

Digital technologies have quickly changed how family members communicate, enjoy themselves, acquire information, and solve daily problems. With the intrusion of the Internet, new digital technologies, and social media into the family environment, the fundamental role of parents has changed (Benedetto & Ingrassia, 2020). Parents are considered the first mediators of children's experiences with digital technologies. According to Padilla-Walker *et al.* (2018), more than half of children's media exposure occurs at home, which presents ample opportunities for parents to get involved in monitoring their children's media use. For children to benefit from accessing digital media and avoid the risks and negative effects of exposure to inappropriate knowledge, parents play a crucial role in their education and development. Literature confirms that, when parents, especially mothers, are confident in having adequate digital skills, they more often intervene (i.e., with rules and reinforcement strategies) with their children (Benedetto & Ingrassia, 2020; Tennakoon *et al.*, 2018; Ilomäki *et al.*, 2014). The present study is underpinned by Potter's cognitive theory of media literacy for assessing digital parental competency (Potter, 2004). Potter (2004) and Lin *et al.* (2013) make the case that media literacy should go beyond only being aware of how media affects society, and they even offer a methodology to do so. Potter's cognitive theory of media literacy introduces three fundamental building elements: knowledge, skills, and personal locus. This model explores how people naturally become media literate, highlighting the cognitive processes that accompany it (Alvermann & Hagood, 2000). Knowledge refers to the information individuals possess about media content, its creators, and its societal impact. Skills involve the ability to analyse, evaluate, and interpret media messages critically. Personal locus refers to the individual's motivation, confidence, and sense of agency in engaging with media content and its broader implications. Within the research framework on digital parenting, writers apply Potter's theory to evaluate parents' ability to successfully navigate the digital world and

https://doi.org/10.18080/jtde.v12n3.969

manage their children's media intake (Mrisho & Dominic, 2023). In general, Potter's cognitive theory of media literacy offers a useful framework for research on digital parenting and sheds light on the intricate interactions that occur between media effects, parental competency, and children's digital wellbeing (Lin *et al.*, 2013; Potter, 2004). Using this paradigm, researchers may assess parents' confidence and motivation in fostering media literacy in their families, as well as their understanding of digital media and their ability to evaluate online information. More specifically, Mrisho & Dominic (2023) provide an assessment instrument to measure parental competency in a digital parenting survey to determine their level of familiarity with digital media platforms, their aptitude for assessing the suitability and reliability of online information, and their level of assurance in overseeing their children's digital activity.

Nonetheless, a wide range of abilities, literacies, and ethical considerations make up digital competence, which is essential for navigating the complexity of the digital environment. Many other theories and frameworks have evolved to conceptualise and convey the characteristics of digital competency as societies move towards more digital surroundings. These theories range from basic ideas like digital literacy to more specialised frameworks like the Diffusion of Innovation theory put forth by Everett Rogers, the Media Literacy Competency Framework by UNESCO, and the European Commission's Digital Competence Framework (DigComp) (Law *et al.*, 2018).

Numerous studies have made significant contributions to our understanding of the dynamics of parental participation, digital competencies, and risk perceptions in the context of digital parenting and its effects on children's wellbeing. Rahayu & Haningsih (2021) conducted a study on the digital competence of informal educators in Indonesia to inform policy formulation. Using the DigCompEdu framework, which assesses informal educators across six areas with 22 questions, the researchers investigated the digital parenting competence of mothers as informal educators. The survey, conducted from December 2019 to January 2020, included both offline and online methods, resulting in 437 valid responses from mothers in the Special Region of Yogyakarta (DIY) province of Indonesia. The mixed method of study found that most mothers act as informal educators at home, yet face limitations due to motherhood priorities, language barriers, and the complexity of technology. Despite difficulties in respondent recruitment due to routine monitoring and survey expansion, the study successfully mapped maternal digital parenting competencies using the DigCompEdu framework. This investigation provides valuable insights into the development of the DigCompEdu framework for informal educators in developing countries, offering a foundation for effective policy formulation and community support. Similarly, Tomczyk (2018) carried out a quantitative study with 260 Polish participants to learn more about parents' digital abilities related to media literacy and online safety. While certain areas, like prevention of cyberbullying, were highlighted as positive abilities, other areas, including copyright protection and safe online interactions, were mentioned as areas of weak competence. However, the emphasis on quantitative indicators sparked worries about ignoring qualitative distinctions and the fundamental causes of digital abilities. Additionally, the research's narrow focus raised issues about equality and how different social circumstances affect digital knowledge.

Meanwhile, Ozerbas & Ocal (2019) used rigorous statistical approaches for data analysis to give insights into how educators, parents, and children perceive digital literacy from the Ankara province. Some interesting trends emerged from their research, including the fact that parents regularly rated their computer proficiency higher than that of their children. Although the study's methodology was strong, its efficacy was hampered by the lack of clear restrictions and practical suggestions. To further highlight the need for more effective teaching methods, the study lacked solutions or treatments to address the issues that were found. A thorough analysis of digital competency in educational research was provided by Ilomäki *et al.* (2014), who also highlighted the difficulties caused by the lack of agreement on terminology and methods. Although the study did an excellent job of identifying the components of digital competence, it might have had a stronger impact by offering more specific recommendations for the creation of policies and practices in education. Similarly, Kallas & Pedaste (2022) investigated how digital competency affected the efficacy of e-learning, stressing the significance of taking both attitudes and skills into account. However, the study has limitations due to its restricted focus on Estonia and its disregard for societal factors.

By creating the Digital Competence Index (DCI) for evaluation purposes, Soldatova & Rasskazova (2014) made a significant contribution to the knowledge of digital competence among Russian adolescents and parents. Although the study was restricted in its generalisability due to its concentration on Russian participants, it provided useful insights into the construction of a complete psychological model and the emphasis on educational programmes required to build digital competence. Meanwhile, Benedetto & Ingrassia (2020) examined the difficulties of digital parenting and mediation techniques, emphasising the advantages of indulgent parenting as well as the drawbacks of strict parenting. However, the study's usefulness was limited due to its concentration on Russian individuals.

On another note, Livingstone & Byrne (2018) discussed the difficulties parents have while using digital parenting across a variety of global socioeconomic brackets, highlighting the requirement for regulations that take these circumstances into account. Similarly, parental views of digital hazards and training needs were the subject of an investigation by İnan-Kaya *et al.* (2018), who provided insights into the difficulties parents encounter when navigating the digital realm. However, issues with generalisability were brought up by restrictions on sample size and specificity. Lastly, digital parenting tactics were the subject of Modecki *et al.* (2022) which also highlighted gaps in qualitative evidence and measurement quality while offering a new framework for assessment.

A review of existing studies reveals that prior research has examined various aspects of digital parenting, including parental competencies, perceptions of online safety, and the impact on children's development. They highlight the importance of understanding digital competence in education, the significance of parent-child relationships, and the need for regulatory measures (Kallas & Pedaste, 2022; Modecki *et al.*, 2022; Benedetto & Ingrassia, 2020; Rahayu & Haningsih, 2021; Ozerbas & Ocal, 2019; Livingstone & Byrne, 2018; Tomczyk, 2018; Ilomäki *et al.*, 2014; Soldatova & Rasskazova, 2014). However, challenges persist regarding terminology, methodology, and generalisability across diverse contexts. Additionally, there is a call for more robust evidence and measurement tools to guide effective digital parenting strategies.

Also, the complex nature of mothers' digital competency in the Indian setting and its impact on their children's digital development has received minimal attention in prior studies. Understanding how maternal digital competence conforms to or defies established cultural norms is critical given India's distinct socio-cultural dynamics, where traditional values collide with fast technological innovation. Furthermore, using digital competence assessment tools is complicated by regional differences in access to computing devices (Martínez-Domínguez & Mora-Rivera, 2020; Law *et al.*, 2018), underscoring the need for measurement instruments that are appropriate for the Indian context. There is also a dearth of knowledge about how a mothers' digital proficiency interacts with Indian cultural norms and expectations to influence children's cognitive development and emotional health. A thorough study carried out in India can shed light on mothers' responsibilities as digital educators, the challenges they encounter, and how their digital competency affects their children's overall digital development. To create focused projects, instructional plans, and policies that take into account India's distinct sociocultural context and add to a complex global conversation on digital competency, bridging this divide is crucial.

Data and Methodology

Research context

A cross-sectional research approach was used in this study to assess and classify mothers according to their levels of digital competency. The study has adopted Modified Digital Competency Education for Informal Educators — modified Dig Comp Edu (Table 1). The questionnaire has been adopted from Rahayu & Haningsih (2021) (Table 2). Mothers (professionals and homemakers) in the study possess a wide range of educational attainments, from primary school to doctorate levels, apart from working in diverse sectors.

Table 1. Modified DigCompEdu Framework

Category	DigCompEdu (Formal Educators)	Modified DigCompEdu for Informal Educators	Competence Description
Competence Area	Professional Engagement	Digital Parenting Resources	Resource selection, resource creation/modification, and resource management/protection
	Digital Resources	Parenting Activities	Parenting, guidance, collaborative learning, self-regulated learning
	Teaching	Empowering Children (Used in this study)	Accessibility, personalization, and active participation
	Assessment	Facilitating Children's Digital Competence (Used in this study)	Information literacy, communication, content creation, responsible use, problem-solving
	Empowering Learners		
	Facilitating Learner's Digital Competence		
Proficiency Level	Newcomer	Novice	Limited familiarity, no use, or inability to operate technology
	Explorer	Basic	Competence in operating with low to moderate frequency
	Integrator	Independent	Proficiency in frequent operations, simple resource modification
	Expert	Proficient	Advanced skills encompassing the ability to evaluate personal use and modify technology
	Leader		
	Pioneer		

Source: Rahayu & Haningsih (2021)

Survey design

Questionnaires were given to 1000 mothers through a convenient sampling method from wards 29 and 30, Kochi Corporation in the Ernakulam District of Kerala. The Digital Competency Framework for Informal Educators 2018, which was released by the European Commission was used to measure digital competency with a primary focus on digital parenting capability. The present study adopted the Modified Digital Competency Education (modified-Dig Comp Edu) Framework for assessing the digital competency of informal educators. Differentiation, Accessibility, Active Participation, Information Literacy, Digital Communication, Digital Content Creation, Responsible Use, and Problem-Solving were among the many criteria that were evaluated (Rahayu & Haningsih, 2021) (Table 1).

https://doi.org/10.18080/jtde.v12n3.969

Table 2. Individual Statements of Each Digital Competence Area

Statements	Items	Items
Accessibility & Inclusion	I don't know if technology is friendly for children with special needs	I am aware that digital technology can improve parenting for children with special needs
	I select digital parenting strategies that adapt children to a digital context	I select digital parenting strategies fitted to children's competence, attitudes, and misuse
Differentiation & Personalization	I don't know how technology helps in personalized parenting opportunities	I am aware that technology activities have different levels of difficulty & speed
	I select & use digital parenting activities that allow children to proceed at different speeds/difficulty	I design a personalized digital parenting plan according to my children's needs
Active Participation	I don't know how technology could encourage children to be motivated	I use technology to explain a new concept to children in a motivating way
	I choose the most appropriate tool for fostering children to be active & motivated	I design the use of technology for parenting according to my children's potential
Information & Media Literacy	I don't know how I foster my children's information and media literacy	I encourage children to use digital technology for information retrieval
	I guide children to find information & assess its reliability from different sources	I reflect on improving my parenting strategies in fostering children's information & media literacy
Digital Communication & Collaboration	I don't know that technology could be used for communication/collaboration	I encourage children to communicate with friends/teachers
	I guide children to select the right communication channels by respecting behavioural norms	I design parenting activities so that children can use digital technology effectively and responsibly
Digital Content Creation	I don't encourage children to make digital content	I encourage children to express themselves using digital technology
	I encourage children to share their digital production by including the copyright	I guide children in designing and publishing complex digital products
Responsible Use	I did not guide children to be aware of the impact of technology	I foster children's awareness of the benefits & drawbacks of technology
	I guide children to protect their privacy & personal data	I enable children to understand risks and threats in digital environments
Digital Problem Solving	I don't know that my children might be able to solve technical problems	I encourage children to solve technical problems using trial & error
	I implement parenting activities in which children use technology creatively	I allow children to investigate the benefits and drawbacks of different technological solutions to a problem

Source: Rahayu & Haningsih (2021)

Users were classified into four competency categories: Novice, Basic, Independent, and Proficient users (Rahayu & Haningsih, 2021). People who are at the 'novice level' show signs of unfamiliarity, lack of usage, or incapacity to utilise technology. Conversely, 'basic users' can

https://doi.org/10.18080/jtde.v12n3.969

urge their children to utilise technology and may operate at a low to moderate frequency. As they advance to the level of 'independent users', people show regular operations, basic resource modification, and the capacity to encourage children to use and share technology. Lastly, 'proficient users' demonstrate sophisticated abilities, such as the capacity to assess their technology usage, make intricate modifications to it, create activities, and teach children how to utilise technology sensibly.

On the other hand, digital competence is defined by the European Commission's Digital Competence Framework (DigComp 2.2) as "the confident and critical use of digital technologies for work, leisure, and communication" (Brande *et al.*, 2018). It is divided into five main categories: problem-solving, digital content creation, communication and collaboration, information and data literacy, and safety. Vendors, independent organizations, and nations offer a variety of digital competence evaluation instruments, most of which are computer-based (Martínez-Domínguez & Mora-Rivera, 2020; Law *et al.*, 2018).

Unfortunately, due to the limited ownership of computing devices, this strategy is challenging to implement in developing countries. Consequently, a practical measurement tool is required, which is the DigCompEdu, a framework designed for formal, informal, and non-formal educators (Rahayu & Haningsih, 2021; Brande *et al.*, 2018). DigCompEdu's goals are to encourage citizens in European Community member nations to gain digital skills and to increase educational innovation (Rahayu & Haningsih, 2021; Brande *et al.*, 2018). By using 22 questions and a progression model divided into six stages, the framework seeks to identify and define the digital competencies of educators. This will enable educators to recognise their current level of competence and subsequently advance it. The six areas that make up the skills are focused on various facets of an educator's professional activities: Professional Engagement, Digital Resources, Teaching and Learning, Assessment, Empowering Learners, and Facilitating Learner's Digital Competence.

The original six competency categories of the Modified DigCompEdu framework for informal educators are expanded to incorporate the critical role that mothers play as digital parents (Rahayu & Haningsih, 2021). Maternal competence is essential in determining a child's growth and wellbeing (Nagy *et al.*, 2023; Balleys, 2022). Mothers' wide range of abilities and skills can influence many facets of their children's lives, including their cognitive growth and emotional wellbeing (Balleys, 2022; Brito *et al.*, 2017). Under this modified framework, mothers select, create, and manage resources to provide a secure and stimulating online environment for their children. Mothers play an active role in fostering their children's digital literacy and competent navigation of the digital world by engaging in parenting activities such as guiding, collaborative learning, and the encouragement of self-regulated learning (Balleys, 2022; Brito *et al.*, 2017). By facilitating accessibility, personalisation, and active engagement,

https://doi.org/10.18080/jtde.v12n3.969

mothers may empower their children by customising digital experiences to meet their specific requirements. As informal educators, mothers play a crucial role in facilitating children's digital competence by fostering information literacy, effective communication, content creation, responsible technology use, and problem-solving skills. Mothers may develop from not being familiar with technology to advanced abilities by applying the four competence levels (Novice, Basic, Independent, and Proficient), which are in line with the larger DigCompEdu structure for formal instructors (Rahayu & Haningsih, 2021). This modified framework emphasises the importance of mothers as major contributors to the larger aims of promoting digitally competent and responsible citizens, acknowledging the significant role that mothers play in forming their children's digital skills within the informal learning environment (Rahayu & Haningsih, 2021).

Statistical analysis

An effective statistical method for grouping people into various groups based on a variety of criteria is discriminant analysis (Wagner & Growe, 2022). Mothers were divided into distinct proficiency groups using discriminant analysis to maximise variation between groups and minimise variance within groups. The analysis was performed using IBM SPSS Version 21.

Results

Demographic profile

Table 3. Demographic Profile

Education (Frequency)	Occupation (Frequency)	Internet Access (Frequency)	Place of Residence (Frequency)
Elementary school (76)	Homemaker (312)	Mobile Internet subscription (614)	Corporation (326)
Pre-degree (148)	Professionals (220)	Home Internet subscription (386)	Municipality (276)
Graduate (453)	IT Professionals (74)		Grama panchayath (398)
Postgraduate (238)	Govt Employee (73)		
PhD (85)	Private sector (262)		
	Engineer (59)		
Total (1000)	1000	1000	1000

Source: Author's work

The demographic profile of mothers in the district reveals a diverse range of characteristics (Table 3). In terms of education, a significant portion is graduates (45.3%), followed by those with postgraduate degrees (23.8%), pre-degree qualifications (14.8%), elementary school education (7.6%), and PhDs (8.5%). Concerning occupation, the largest group consists of homemakers (31.2%), with corporate employees (32.6%) forming the largest professional segment, followed by professionals (22%), IT professionals (7.4%), government employees (7.3%), private sector workers (26.2%), and engineers (5.9%). Regarding Internet access, the majority use mobile Internet subscriptions (61.4%), while the rest have home Internet subscriptions (38.6%). The place of residence is varied, with the largest group living in villages (39.8%), followed by those in municipalities (27.6%) and corporations (32.6%). This data showcases the educational diversity, professional engagement, Internet usage, and residential distribution of mothers in Ernakulam district, highlighting a balanced mix of rural and urban settings.

Discriminant analysis

The present study considers the concept of digital competence of informal educators, where the framework has two components: a sociodemographic component and a piece on digital competence that was based on a modified DigCompEdu framework (Rahayu & Haningsih, 2021). The modified version of digital competence for informal educators has four competence areas: Digital Parenting Resources; Parenting Activities; Empowering Children; and Facilitating Children's Digital Competence. These areas are divided into several items that assess four proficiency levels: Novice, Basic, Independent, and Proficient User Levels (Rahayu & Haningsih, 2021).

Discriminant analysis is used to identify the variables that can distinguish between two or more naturally occurring categories (Walde, 2014). The goal of this analysis is to identify the variables that significantly contribute to discrimination among various groups that represent different levels of competency (Novice, Basic, Independent, and Proficient), and which variables, encompassing eight proficiency areas (Accessibility, Differentiation, Active Participation, Information Literacy, Digital Communication, Digital Content Creation, Responsible Use, and Problem-Solving), do so. The ability of these factors to successfully discriminate between these proficiency groups is often a deciding factor in their selection (Walde, 2014). The discriminant function, usually referred to as the linear combination for discriminant analysis, is generated from the following equation:

$$Z_{ik} = \beta_{0i} + \beta_{1i} X_{1k} + \cdots + \beta_{ji} X_{jk} \quad (1)$$

where Xj_k is the independent variable j (j = 1, 2,..., J) for object k and Z_{ik} is the discriminant score of the discriminant function i (i = 1, 2,..., G1). The discriminant weight for independent variable j is represented by the symbol ji function i, and oi is the discriminant function i's constant (Wagner & Growe, 2022; Walde, 2014).

https://doi.org/10.18080/jtde.v12n3.969

Wilks' Lambda is a multivariate test statistic that measures the extent of difference between groups that affect the effect size of the difference (Wagner & Growe, 2022). Greater group separation is indicated by smaller Wilks' Lambda values (Table 4).

Table 4. Tests of Equality of Group Means (ANOVA Table)

	Wilks' Lambda	F	df1	df2	Sig.
Accessibility	.148	1904.141	3	996	.000
Differentiation	.639	187.381	3	996	.000
Active Participation	.406	486.535	3	996	.000
Information Literacy	.730	122.589	3	996	.000
Digital Communication	.607	214.673	3	996	.000
Digital Content Creation	.773	97.295	3	996	.000
Responsible Use	.573	247.832	3	996	.000
Problem Solving	.782	92.617	3	996	.000

Source: Data Analysis

It is noticed that there are significant variations between groups for each of the dependent variables, as indicated by the fact that all of the values in the result are fairly small. Each Wilks' Lambda value has a corresponding F-statistic. A greater influence is indicated by larger F-values (Wagner & Growe, 2022). The F-values all appear to be rather high in this case. The p-value for each F-test is listed in the Sig. column. It shows the likelihood of obtaining the observed F-statistic in the absence of any significant group differences for the dependent variable. There are substantial differences between groups for each dependent variable, as shown by the analysis's p-values, which are all very close to zero (Wagner & Growe, 2022).

The outcomes of Box's Test of Equality of Covariance Matrices (Table 5), a statistical method used in discriminant analysis, are shown in Table 5 and Table 6.

Table 5. Box's Test of Equality of Covariance Matrices

Proficiency Level	Rank	Log Determinant
Novice	1	.[a]
Basic	8	9.418
Independent	8	8.803
Proficient	8	-3.783
Pooled within-groups	8	9.385

a. Singular
Source: Data Analysis

Table 6. Test Results

Box's M		2541.245
F	Approx.	34.758
	df1	72
	df2	840567.124
	Sig.	.000

Tests null hypothesis of equal population covariance matrices.
Source: Data Analysis

This test evaluates the equality of the covariance matrices across several groups (Wagner & Growe, 2022). It assists in determining whether common covariance matrices or group-specific ones should be utilised. For each proficiency level group, the table provides the rank and natural logarithms of the covariance matrix determinants, in addition to the pooled within-groups covariance matrix. The Basic, Independent, and Proficient groups all have ranks of 8, which corresponds to eight dimensions or variables. Natural logarithms of the covariance matrix determinants are contained in the Log Determinant column, providing information on the distribution and variability of the data. To assess the importance of fluctuations in the covariance matrix, an approximation F-statistic (F (Approx.)) is linked to the test statistic, Box's M, which is 2541.245 (Table 6). The hypothesis of equal population covariance matrices is rejected by the low p-value (e.g., .000), which implies significant differences across group covariance matrices (Wagner & Growe, 2022).

Summary of canonical discriminant functions

The outcomes of canonical discriminant functions, classification, and related statistics are presented in the following sections. The amount of variation that each canonical discriminant function explains is represented by its eigenvalues (Table 7).

Table 7. Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	6.208[a]	94.0	94.0	.928
2	.283[a]	4.3	98.3	.470
3	.113[a]	1.7	100.0	.319

a. First 3 canonical discriminant functions were used in the analysis
Source: Data Analysis

Three discriminant functions are displayed in Table 7. Function 1 accounts for 94% of the variance, with Function 2 accounting for 4.3% and Function 3 for 1.7%. These eigenvalues assist in understanding the relative significance of each function in group distinction (Wagner & Growe, 2022).

Results for tests 1 through 3, 2 through 3, and Function 3 are displayed in Wilks' Lambda table (Table 8).

Table 8. Wilks'Lambda

Test of Function(s)	Wilks' Lambda	Chi-squared	df	Sig.
1 through 3	.097	2315.887	24	.000
2 through 3	.700	354.563	14	.000
3	.898	106.727	6	.000

Source: Data Analysis

Wilks' Lambda has values between 0 and 1. Greater discrimination is indicated by smaller values (Wagner & Growe, 2022). Wilks' Lambda is 0.097 for the first test (1 through 3), suggesting a significant group separation.

The unstandardised discriminant function coefficients from Table 9 are utilised to build the actual prediction equation for categorising new cases.

Table 9. Standardised Canonical Discriminant Function Coefficients

	Function		
	1	2	3
Accessibility	.727	-.165	-.091
Differentiation	-.059	.321	.307
Active Participation	.009	-.082	-.002
Information Literacy	-.003	-.207	.082
Digital Communication	-.011	.368	-.412
Digital Content Creation	-.036	-.050	-.063
Responsible Use	.166	.174	.111
Problem Solving	-.036	-.193	.254
(Constant)	-11.220	-1.964	-2.242

Unstandardised coefficients
Source: Data Analysis

Equation (2) gives the model created for this study based on the coefficients in Table 9.

$$
\begin{aligned}
Z = -11.220 &+ .727*Accessibility - .059*Differentiation + .009*Active\_Participation \\
&- .003*Information\_Literacy - .011*Digital\_Communication \\
&- .036*Digital\_Content\_Creation + .166*Responsible\_Use \\
&- .036*Problem\_Solving
\end{aligned}
\tag{2}
$$

The structure matrix in Table 10 provides the correlations between the original variables and the discriminant functions. For a certain function, variables with greater absolute correlation values are more crucial for differentiating between groups (Wagner & Growe, 2022).

Table 10. Structure Matrix

	Function		
	1	2	3
Accessibility	.946*	.098	.050
Active Participation	.457*	-.183	.283
Digital Content Creation	.210*	.117	-.047
Digital Communication	.254	.486*	-.184
Responsible Use	.325	.401*	.276
Problem Solving	.121	-.224*	.075
Differentiation	.160	.614	.647*
Information Literacy	.251	.105	.562*

Largest absolute correlation between each variable and any discriminant function
Source: Data Analysis

Functions at Group Centroids (Table 11) indicate the canonical discriminant functions that are not standardised and are assessed at group means (Wagner & Growe, 2022). The result in Table 11 displays where each group is located inside the discriminant space.

Table 11. Functions at Group Centroids

Proficiency Level	Function		
	1	2	3
Novice	-7.027	-1.254	2.153
Basic	-3.322	-.347	-.332
Independent	.458	.433	.076
Proficient	3.851	-.831	-.016

Unstandardised canonical discriminant functions evaluated at group means
Source: Data Analysis

Classification statistics

Classification Function- Based on their variable values, new cases are divided into groups using coefficients (Table 12). The coefficient of each variable helps to sort the instances into the categories that were expected.

Table 12. Classification Function Coefficients

	Proficiency Level			
	Novice	Basic	Independent	Proficient
Accessibility	1.969	4.739	7.321	10.004
Differentiation	1.156	.466	.621	-.012
Active Participation	-.349	-.383	-.413	-.278
Information Literacy	.326	-.076	-.217	.027
Digital Communication	-.970	.347	.423	-.042
Digital Content Creation	.071	.049	-.151	-.203
Responsible Use	1.249	1.745	2.553	2.886
Problem Solving	1.805	.866	.682	.782
(Constant)	-27.062	-42.688	-82.112	-125.062

Fisher's linear discriminant functions
Source: Data Analysis

Table 13. Classification Results

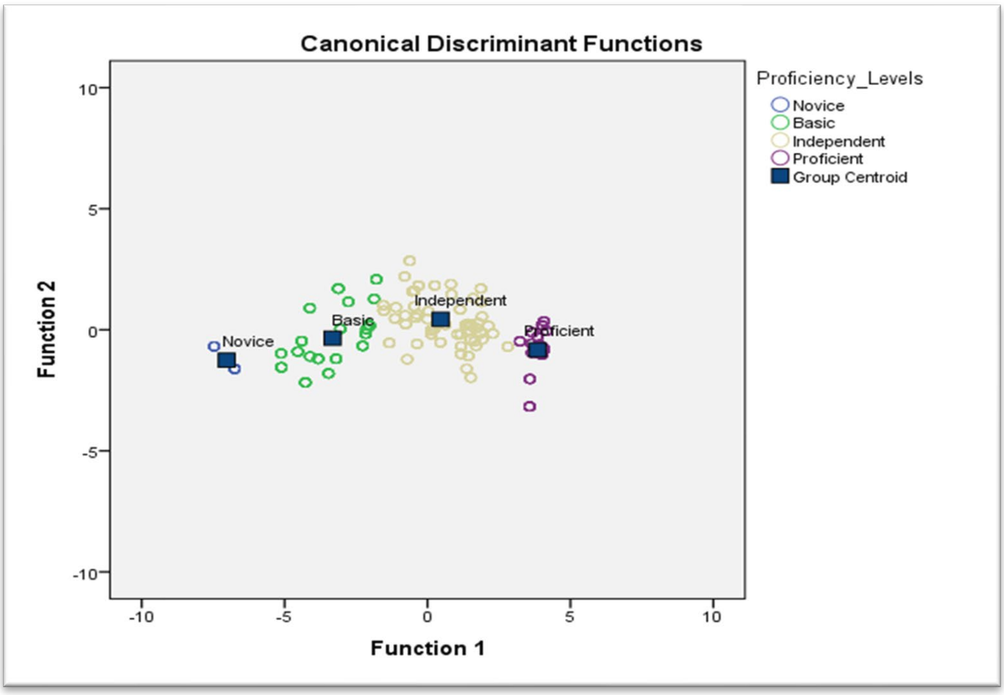
		Proficiency Level	Predicted Group Membership				Total
			Novice	Basic	Independent	Proficient	
Original	Count	Novice	18	0	0	0	18
		Basic	0	224	15	0	239
		Independent	0	15	536	21	572
		Proficient	0	0	0	171	171
	%	Novice	100.0	.0	.0	.0	100.0
		Basic	.0	93.7	6.3	.0	100.0
		Independent	.0	2.6	93.7	3.7	100.0
		Proficient	.0	.0	.0	100.0	100.0

94.9% of original grouped cases correctly classified.
Source: Data Analysis

https://doi.org/10.18080/jtde.v12n3.969

Table 13 displays the quantity and percentage of cases that were appropriately categorised into each group according to proficiency level; 93.7% of the initial cases were appropriately categorised into their real proficiency levels in this instance.

The scatter plot used in our study (Figure 1) graphically shows data points in a multidimensional space. The axes and group centroids, which display the average position of each group, are defined by canonical discriminant functions (Wagner & Growe, 2022). In the plot, distinct symbols or colours are used to denote different groupings. It is possible to evaluate the performance of the discriminant functions by examining the scatter plot, data distribution, and group centroids. The graph clearly shows effective group separation, and the distances between centroids indicate whether the variables and groupings are similar or dissimilar.



Source: Data Analysis

Figure 1. Group Centroids and Scatter Plot of the Canonical Discriminant Function

Discussion

Significant insights into the relationships between digital parenting competences and competency levels are offered by the findings. The initial canonical discriminant function demonstrates a significant explanatory power, accounting for a significant amount (95.8%) of the observed variance, supporting its strong discriminative capacity across the proficiency level, according to the eigenvalues study (Obudho *et al.*, 2022; Wagner & Growe, 2022). The separation of proficiency level in the multi-dimensional space described by the functions is visually shown by the scatter plot showing group centroid and canonical discriminant

https://doi.org/10.18080/jtde.v12n3.969

functions (see Figure 1). The Novice and Proficient groups display clear separation in this plot, which is notable as it demonstrates effective group discrimination. Additionally, the classification statistics (see Table 13) demonstrate a high percentage of accurate categorisation. This demonstrates the effectiveness of the discriminant functions generated from the study in categorising new cases into their appropriate competency levels based on digital parenting capabilities.

The discriminant analysis classified mothers into four categories based on their digital Proficiency Level: Novice, Basic, Independent, and Proficient. The highest concentration is on Level 3, Independent proficiency, which implies that mothers have good digital competence in managing children's digital technology use. These categories provide insights into the varying digital parenting abilities of mothers and can assist in understanding the impact of digital technology on parenting.

The important elements that differentiate mothers' degrees of digital proficiency are shown by the structural matrix (Table 10). The factors that have the greatest significance for Function 1 are digital content creation (.210), active participation (.457), and accessibility (.946). This implies that the capacity to actively interact with digital technology, produce digital material, and have access to digital tools and resources are crucial factors in differentiating mothers in this role. The primary determinants for Function 2 are digital communication (.486) and responsible usage (.401), showing that the function requires expertise in digital communication as well as responsible use of digital technology, which includes knowledge of online safety and privacy. Even though they are negatively associated, problem-solving abilities (-.224) also matter. Differentiation (.647) and information literacy (.562) are the primary factors for Function 3, emphasising the significance of the aptitude to locate, assess, and apply knowledge in addition to the ability to recognise and successfully use various digital tools and platforms. The canonical discriminant function coefficients (Table 9) delineate critical variables impacting Keralan women's degrees of digital competency, specifically about digital parenting. Function 1 of accessibility highlights the necessity of enhancing access to digital tools and information. Function 1 emphasises the importance of active involvement and the creation of digital content, recommending workshops and centres for the development of digital engagement and content creation skills. Function 2 promotes initiatives for online safety education while highlighting digital communication and appropriate use. Function 3 emphasises information literacy and distinction, including resources for assessing digital material as well as advanced literacy training. The aforementioned programmes are designed to enable mothers to successfully and ethically guide their children's digital activities.

Through the modification of the DigCompEdu framework, the study aims to create a tool specifically tailored to assess maternal competence as an informal educator in the home setting (Rahayu & Haningsih, 2021). This underscores a comprehensive approach that acknowledges the significance of both technological knowledge and effective pedagogical practices in the realm of digital parenting. Women must possess fundamental digital skills to contribute positively to society, the workplace, and the family (Huang *et al.*, 2018). In light of mothers' responsibility as educators for their school-age children, the usage of modified-DigCompEdu enriches the viewpoint on digital parenting (Rahayu & Haningsih, 2021).

Furthermore, the acknowledgement of sociodemographic factors influencing digital skills parallels Juhász *et al.*'s (2022) examination, reinforcing the need for interventions tailored to specific contexts. Additionally, Sulistyo's (2022) emphasis on parental involvement in digital literacy concurs with the recognition of digital parenting competences in the current analysis. Also, Ilomäki *et al.*'s (2014) acknowledgement of a lack of standardised concepts in digital competence aligns with the current study's emphasis on the need for cohesive definitions and frameworks. Finally, the importance of behavioural intention in digital competence, as highlighted in Kallas & Pedaste (2022), resonates with the present analysis, emphasising the significance of strategies for enhancing digital competencies. In sum, the current study not only reaffirms but also enriches the existing literature by contributing nuanced insights into the intricate relationships between digital parenting competences and competency levels.

Theoretical contributions

The application of media literacy theory, namely Potter's framework, to the setting of digital parenting is the study's theoretical contribution. According to Potter's conceptualisation of media literacy theory, media messages should be critically analysed and understood, and skills for efficiently navigating and interpreting media information should be developed. This theoretical viewpoint broadens our understanding of how parents may navigate the digital world and encourage their children to behave responsibly and safely online by bringing it into the study of digital parenting.

Specifically, the study applies Potter's framework to examine mothers' competency levels in digital parenting, encompassing their ability to critically evaluate and manage their children's digital media use. By utilising media literacy theory as a lens through which to analyse parental practices and behaviours in the digital realm, the study enhances our understanding of the complex interplay between media influences, parental guidance, and children's digital wellbeing.

Another theoretical contribution of this study lies in its adaptation of media literacy principles to the domain of digital parenting, shedding light on the strategies and skills that mothers employ to navigate the digital landscape effectively. By drawing on Potter's framework, the study provides valuable insights into how parents can promote media literacy skills and responsible digital citizenship among their children, contributing to ongoing discussions surrounding digital parenting practices and interventions. Overall, the study's theoretical contribution underscores the importance of integrating media literacy theory into the study of digital parenting to enhance parental efficacy and promote positive digital experiences for children.

Practical implications

Practical implications of the work are evident in several domains. Concerning educational institutions, the results can support developing specialised educational initiatives, while educators may use the information to improve the practices of educational institutions. These projects can entail changing the current curriculum to incorporate parents in the instruction of digital citizenship and responsible behaviour on the Internet. To successfully address the present digital issues, schools and educational institutions should also conduct seminars targeted at enhancing parents' digital literacy and talents. This will create more collaboration between schools and families.

In addition, IT firms and app developers may play an important role in encouraging digital parenting. These businesses can accommodate mothers with different levels of digital proficiency by creating digital tools and platforms with user-friendly features, which makes it easier for parents to navigate and supervise their children's online activity. Improving the parental control capabilities of software programs might help mothers even more in their endeavours to protect and care for their children online.

By establishing guidelines for information that is appropriate for children's age and offering resources for monitoring and screening their online activities, media and content makers may support digital parenting. Media companies can help families navigate the intricacies of the digital ecosystem while protecting children from harmful content and online hazards by providing instructive and entertaining resources that teach mothers how to be better digital parents.

The results of this study can help shape legislation and initiatives targeted at curbing children's rampant use of digital gadgets and lessening their harmful impacts. Tailored solutions may be developed by considering the mothers' different degrees of digital proficiency. Multiple interfaces may be used to create user-friendly parental control apps like basic dashboards for basic monitoring in beginner-friendly versions, and more complex features like customisable filters and comprehensive use statistics in advanced versions. To improve mothers' understanding, educational materials may be integrated, such as interactive modules on current digital trends, responsible Internet usage, and digital safety. Parents can be alerted about potentially dangerous activities in real time and given guidance on proper measures to

https://doi.org/10.18080/jtde.v12n3.969

take. Family activity trackers may also assist in striking a balance between digital and offline activities by tracking screen time and recommending offline activities based on the family's online and offline use patterns.

The findings of this study have important ramifications for developing mother-focused digital literacy initiatives. Targeted campaigns may be developed by identifying different degrees of digital competency. This will help mothers become more adept at utilising technology to better protect and mentor their children online. Programmes for novice and basic users should concentrate on teaching fundamental digital skills, such as how to use computers and smartphones and the fundamentals of digital safety. Search engines, email accounts, and the Internet can all be taught through interactive courses. Essential support can also be obtained through simple-to-follow printed and digital guides, peer support groups for community members to share experiences with, and one-on-one tutoring sessions for specific questions.

Programmes should improve online safety, basic troubleshooting, and digital communication skills for independent users. Independent users can improve their skills with interactive online courses on social media responsibility and privacy management, webinars with experts in digital literacy for deeper insights, parent-child digital activities like creating a family blog or learning about digital citizenship together, and resource libraries with interactive articles, videos, and content. Programmes should address advanced digital tools, apps, and cybersecurity for competent users. Proficient users can be empowered through advanced training sessions on subjects like creating digital content, cybersecurity, and educational apps; mentoring programmes, where more seasoned mothers help less seasoned ones; comprehensive plans for digital literacy for families that include goals and tactics for both parents and children; and involvement in policy-making to support community initiatives on digital literacy and cyber safety. With this targeted strategy, mothers may better manage contemporary technology and better guide and safeguard their children in the digital age.

Lastly, academicians and researchers play a critical role in expanding our knowledge of digital parenting and how it affects the health of children. The study provides a useful foundation for future investigations into how mothers' digital proficiency is changing and how that can affect their children's digital wellbeing. Academic institutions may play a crucial role in preparing individuals with the information and skills required to support families in the digital era by offering courses or certificates designed to educate professionals in helping mothers with the demands of digital parenting.

Policy implications

The research results have several significant policy ramifications, especially for child safety, technology control, and education. Governments and institutions can foster a supportive environment that improves digital parenting practices and ultimately contributes to the digital wellbeing of children in Kerala by incorporating these recommendations into currently implemented policies, such as the National Digital Literacy Mission (NDLM) and the Digital India initiative.

1. Educational Policy: In line with the objectives of the National Digital Literacy Mission, 2015 policymakers should incorporate parental education and digital literacy into school curriculums. Community centres and educational institutions can implement programmes that educate parents, particularly mothers, on digital parenting techniques. These courses should strongly emphasise fostering healthy online conduct, controlling children's screen time, and critically evaluating digital information. Parents will be better prepared to mentor their children in the digital era by including these programmes in the curriculum.
2. Technology Regulation: To ensure that digital devices and applications have strong parental control capabilities, government authorities should work with technology companies to draft rules. User-friendly parental control features that enable parents to properly monitor and manage their children's online activities might be mandated by policy. Increased restrictions on harmful and inappropriate content should be implemented, along with standards for the age-appropriateness of content that is available to minors. The Indian Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, should be complied with through these procedures.
3. Child Protection: It is necessary to reinforce laws intended to safeguard children online. Campaigns that raise awareness of Internet dangers, including cyberbullying, privacy violations, and exposure to offensive material, fall under this category. Clear recommendations on the obligations of parents, educators, and technology providers in protecting children's online experiences should be incorporated into current legal frameworks. Funds should be set aside for support programmes that offer guidance and intervention in situations of addiction and digital abuse. These initiatives will strengthen children's digital safety net under the Indian Protection of Children from Sexual Offences (POSCO) Act.

By incorporating these suggestions into current legislation, India may promote a more welcoming and helpful digital environment, guaranteeing children's digital welfare and equipping parents with the skills and information they need.

https://doi.org/10.18080/jtde.v12n3.969

Conclusion, Limitations, and Future Research

In conclusion, the field of digital parenting is a promising and developing one, especially in the context of India. The findings of our research shed light on the situation in Kerala, where mothers demonstrate a basic understanding of the digital landscape and skilfully control their children's use of digital media. Recognising that digital proficiency is an important skill that should be developed from a young age is vital. This skill can be developed through parenting, which is a crucial method. It is the responsibility of mothers to acquire the skills needed to help their children take advantage of the countless opportunities offered by the digital world while also understanding the possible threats that may exist there. Therefore, understanding mothers' levels of proficiency is a crucial first step in developing effective digital parenting practices. Mothers should make an effort to promote their children's overall growth as their primary guardians by finding a balance between responsible use of digital media and participation in physical activity. This balanced strategy can make sure children are ready to succeed in a world that is becoming more and more digital, while also leading a healthy and balanced lifestyle. Using the knowledge gained, customised educational plans, focused interventions, or support initiatives can be created to tackle particular problems or make the most of the particular advantages that each proficiency group possesses. The effectiveness of activities aiming at enhancing digital parenting capabilities at various levels of proficiency may be improved by this personalised approach.

The study's findings may be limited by the representativeness of the sample, potentially introducing sample bias. If the sample is not diverse enough or does not accurately reflect the broader population, the generalisability of the results to all mothers in Kerala, particularly in the Ernakulam District, may be compromised. Additionally, reliance on self-reported data may introduce bias and inaccuracies. Mothers might overstate or understate their digital competence levels or parenting practices, impacting the reliability of the findings. Furthermore, the study might not capture the long-term effects of digital parenting strategies on children's wellbeing. Longitudinal studies would be required to assess the sustained impact of maternal digital competence on children as they grow and interact with evolving digital environments.

In the pursuit of advancing our understanding of digital parenting in Kerala, particularly in Ernakulam District, several promising avenues for future research and initiatives emerge. First and foremost, conducting longitudinal studies to track the digital competence levels of mothers and their impact on children's wellbeing over an extended period represents a critical next step. This approach allows researchers to unravel the evolving dynamics of digital parenting practices and their sustained effects as children grow in a rapidly changing technological landscape.

Additionally, delving into more in-depth cultural analyses offers the potential to uncover subtle nuances and traditions that shape digital parenting behaviours. Qualitative research methods could be employed to capture the rich context and intricacies associated with digital parenting in this region. Comparative studies across various districts or states within Kerala would be valuable in identifying regional variations and factors influencing digital competence. Implementing and rigorously evaluating targeted interventions based on the identified digital competence levels could provide evidence-based insights into the effectiveness of specific programs or policies aimed at enhancing digital parenting skills. Furthermore, staying attuned to technological advances and their implications for parenting strategies is crucial, as ongoing research will need to adapt and evolve to address emerging challenges and opportunities in the digital landscape.

References

Alvermann, D. E., & Hagood, M. C. (2000). Critical media literacy: Research, theory, and practice in "new times". *Journal of Educational Research*, *93*(3), 193–205. https://doi.org/10.1080/00220670009598707

Balleys, C. (2022). Familial digital mediation as a gendered issue between parents. *Media Culture & Society*, *44*(8), 1559–1575. https://doi.org/10.1177/01634437221119020

Benedetto, L., & Ingrassia, M. (2020). Digital parenting: Raising and protecting Children in Media World. In Parent. [Working Title]. *IntechOpen*. https://www.intechopen.com/online-first/digital-parenting-raisingand-protecting-children-in-media-world.

Biswas, S., & Narendranath. (2015). Kudumbashree-PRI Collaboration in Kerala: An Approach to Poverty Alleviation. In *NewsReach*. Wileys.

Brande, L., Carretero, S., Vuorikari, R., & Punie, Y. (2018). DigComp 2.0 – The digital competence framework for citizens. European Commission: Joint Research Centre, Publications Office. https://doi.org/10.2791/11517

Brito, A. D., Jacon, A., Queiroz, J., & Valeriano, D. (2017). Mapping the main vegetation types of Cerrado biome in the year 2000, link to GIS files [Dataset]. PANGAEA. https://doi.org/10.1594/PANGAEA.882605

Carvalho, J., Francisco, R., & Relvas, A. P. (2015). Family functioning and information and communication technologies: How do they relate? A literature review. *Computers in Human Behaviour*, *45*, 99–108. https://doi.org/10.1016/j.chb.2014.11.037

Chacko, E. (2003). Marriage, development, and the status of women in Kerala, India. *Gender & Development*, *11*, 52–59.

Chandrasekaran, N., & Purushothaman, R. (2022). *Bridgital Nation: Solving Technology's People Problem*. Penguin Books Limited. https://books.google.co.in/books?id=yTbKDwAAQBAJ

Chassiakos, Y. R., Radesky, J., Dimitri, C., Moreno, M. A., & Cross, C. (2016). Children and adolescents and digital media. In *American Academy of Pediatrics eBooks* (pp. 567–584). https://doi.org/10.1542/9781610020190-part05-children03

Cingel, D. P., & Krcmar, M. (2013). Predicting Media Use in Very Young Children: The Role of Demographics and Parent Attitudes. *Communication Studies*, *64*(4), 374–394. https://doi.org/10.1080/10510974.2013.770408

Gómez, P., Harris, S. K., Barreiro, C., Isorna, M., & Rial, A. (2017). Profiles of Internet use and parental involvement, and rates of online risks and problematic Internet use among Spanish adolescents. *Computers in Human Behaviour*, *75*, 826–833. https://doi.org/10.1016/j.chb.2017.06.027

Gür, D., & Türel, Y. K. (2022). Parenting in the digital age: Attitudes, controls and limitations regarding children's use of ICT. *Computers & Education*, *183*, 104504. https://doi.org/10.1016/j.compedu.2022.104504

Huang, G., Xiaoqian, L., Chen, W., & Straubhaar, J. D. (2018). Fall-behind parents? The influential factors on digital parenting self-efficacy in disadvantaged communities. *American Behavioural Scientist*, *62*, 1186–1206. http://dx.doi.org/10.1177/0002764218773820.

Ilomäki, L., Paavola, S., Lakkala, M., & Kantosalo, A. (2014). Digital competence – an emergent boundary concept for policy and educational research. *Education and Information Technologies, 21*, 655–679. https://doi.org/10.1007/s10639-014-9346-4

İnan-Kaya, G., Mutlu-Bayraktar, D., & Yılmaz, Z. (2018). Digital Parenting: Perceptions on Digital Risks. *Kalem Uluslararasi Egitim Ve Insan Bilimleri Dergisi*, *14*(1), 137–163. https://doi.org/10.23863/kalem.2018.96

Juhász, T., Kálmán, B. G., Tóth, A., & Horvath, A. (2022). Digital competence development in a few countries of the European Union. *Management & Marketing. Challenges for the Knowledge Society*, *17*, 178–192. https://doi.org/10.2478/mmcks-2022-0010

Kallarakal, A., & Gonsalvez, J. (2021). Perception of Parents and Children on Parenting in Mala, Kerala, India. *European Scientific Journal, ESJ*, *17(33)*, 81. https://doi.org/10.19044/esj.2021.v17n33p81

Kallas, K., & Pedaste, M. (2022). How to Improve the Digital Competence for E-Learning? *Applied Sciences*, *12*(13), 6582. https://doi.org/10.3390/app12136582

Law, N., Woo, D., De La Torre, J., & Wong, K. (2018). *A Global Framework of Reference on Digital Literacy Skills for Indicator 4.4.2*. UNESCO Institute for Statistics, Information Paper No.51. https://hub.hku.hk/bitstream/10722/262055/1/Content.pdf

Lin, T.-B., Li, J.-Y., Deng, F., & Lee, L. (2013). Understanding new media literacy: An explorative theoretical framework. *Educational Technology and Society*, *16*(4), 160–170. Available at https://www.jstor.org/stable/jeductechsoci.16.4.160

Livingstone, S., & Byrne, J. (2018). Parenting in the Digital Age. In *Digital Parenting. The Challenges for Families in the Digital Age* (pp. 19–30). Nordicom. https://norden.diva-portal.org/smash/get/diva2:1535895/FULLTEXT01.pdf

Martínez-Domínguez, M., & Mora-Rivera, J. (2020). Internet adoption and usage patterns in rural Mexico. *Technology in Society*, *60*, 101226. http://dx.doi.org/10.1016/j.techsoc.2019.101226.

Mascheroni, G., & Ólafsson, K. (2014). *Net Children Go Mobile: risks and opportunities*. Second Edition. Milano: Educatt. https://netchildrengomobile.eu/ncgm/wp-content/uploads/2013/07/DEF_NCGM_SecondEdition_Report.pdf

Modecki, K. L., Goldberg, R. E., Wisniewski, P., & Orben, A. (2022). What Is Digital Parenting? A Systematic Review of Past Measurement and Blueprint for the Future. *Perspectives on Psychological Science*, *17*(6), 1673–1691. https://doi.org/10.1177/17456916211072458

Mohan, D., Bashingwa, J. J. H., Tiffin, N., Dhar, D., Mulder, N., George, A., & LeFevre, A. E. (2020). Does having a mobile phone matter? Linking phone access among women to health in India: An exploratory analysis of the National Family Health Survey. *PLoS ONE*, *15*(7), e0236078. https://doi.org/10.1371/journal.pone.0236078

Mrisho, D. H. & Dominic, N. A. (2023). Media Literacy: Concept, Theoretical Explanation, and its Importance in the Digital Age. *East African Journal of Arts and Social Sciences*, *6*(1), 78–85. https://doi.org/10.37284/eajass.6.1.1087

Nagy, B., Czigler, I., Csizmadia, P., File, D., Fáy, N., & Gaál, Z. A. (2023). Investigating the involvement of cognitive control processes in innovative and adaptive creativity and their age-related changes. *Frontiers in Human Neuroscience*, *17*. https://doi.org/10.3389/fnhum.2023.1033508

Nair, M. K. C., Sumaraj, L., Padmamohan, L., Radhakrishnan, R., Nair, V. R., George, B., & Kumar, G. S. (2007). Parenting practices in Kerala: A cross-sectional study. *Vulnerable Children and Youth Studies*, *2*(1), 71–79. https://doi.org/10.1080/17450120601122646

Obudho, M. G., Orwa, G. O., Otieno, R. O., & Were, F. A. (2022). Robust Classification through a Nonparametric Kernel Discriminant Analysis. *Open Journal of Statistics*, *12*, 443–455. https://doi.org/10.4236/ojs.2022.124028

Ozerbas, M. A., & Ocal, F. N. (2019). Digital Literacy Competence Perceptions of Classroom Teachers and Parents Regarding Themselves and Parents' Own Children. *Universal Journal of Educational Research*, *7*(5), 1255–1264. https://doi.org/10.13189/ujer.2019.070511

Padilla-Walker, L. M., Carlo, G., & Memmott-Elison, M. K. (2018). Longitudinal change in adolescents' prosocial behaviour toward strangers, friends, and family. *Journal of Research on Adolescence*, *28*(3), 698–710. https://doi.org/10.1111/jora.12362.

Potter, W. J. (2004). Argument for the need for a cognitive theory of media literacy. *American Behavioural Scientist*, *48*(2), 266–272. https://doi.org/10.1177/0002764204267274

Rahayu, N. W., & Haningsih, S. (2021, September). Digital parenting competence of mother as informal educator is not inline with Internet access. *International Journal of Child-Computer Interaction*, *29*, 100291. https://doi.org/10.1016/j.ijcci.2021.100291

Sela, Y., Zach, M., Amichay-Hamburger, Y., Mishali, M., & Omer, H. (2020). Family environment and problematic Internet use among adolescents: The mediating roles of depression and Fear of Missing Out. *Computers in Human Behaviour*, *106*, 106226. https://doi.org/10.1016/j.chb.2019.106226

Soldatova, G., & Rasskazova, E. (2014). Assessment of the digital competence in Russian adolescents and parents: Digital Competence Index. *Psychology in Russia: State of the Art*, *7*(4), 65–74. https://doi.org/10.11621/pir.2014.0406

Sulistyo, P. B. (2022). Digital Literacy Competence of Parents in Supervising Their Children Using Digital Media. *International Journal of Social Science and Human Research*, *05*(02). https://doi.org/10.47191/ijsshr/v5-i2-32

Tennakoon, H., Saridakis, G., & Mohammed, A.-M. (2018). Child online safety and parental intervention: a study of Sri Lankan Internet users. *Information Technology & People*, *31*, 770–790. https://dx.doi.org/10.1108/ITP-09-2016-0213

Tomczyk, U. (2018). Digital competences of parents in the matter of electronic threats. *SHS Web of Conferences*, *48*, 01004. https://doi.org/10.1051/shsconf/20184801004

Vale, A., Pereira, F., Gonçalves, M., & Matos, M. (2018). Cyber-aggression in adolescence and Internet parenting styles: A study with victims, perpetrators and victim-perpetrators. *Children and Youth Services Review*, *93*, 88–99. https://doi.org/10.1016/j.childyouth.2018.06.021

Wagner, M., & Growe, A. (2022). Patterns of knowledge bases in large city regions in Germany: comparison of cores and their surrounding areas. *Geografiska Annaler: Series B, Human Geography*, *105*(3), 284–304. https://doi.org/10.1080/04353684.2022.2141131

Walde, J. (2014). Discriminant Analysis. In *Lecture Notes*. *Department of Statistics*. University of Innsbruck. Available at: https://www.uibk.ac.at/statistics/personal/janettewalde/lehre/phd_biology/diskriminanzanalyse.pdf

Yazon, A. D., Ang-Manaig, K., Buama, C. A. C., & Tesoro, J. F. B. (2019). Digital literacy, digital competence and research productivity of Educators. *Universal Journal of Educational Research*, *7*, 1734–1743. https://dx.doi.org/10.13189/ujer.2019.070812

https://doi.org/10.18080/jtde.v12n3.969

How Does the Digital Economy Influence the Pursuit of Sustainable Development Goals?

Wajdi Moussa
Higher Institute of Management of Tunis

Nidhal Mgadmi
Faculty of Economic Sciences and Management of Mahdia

Afef Guachaoui
Higher School of Commerce of Tunis

Rym Regaieg
Higher Institute of Management of Tunis

Abstract: Our study aims to examine the impact of the digital economy on sustainable development during international financial crises, the COVID-19 pandemic, and the current war between Russia and Ukraine. To do this, we collected an annual dataset from 1990 to 2022 for twenty-eight developed and twenty-five developing countries, to identify links between environmental, socio-economic, and proxy indicators of the digital economy. We estimate these links using the Within and GLS methods and use the Hausman (1978) test for individual effects to determine the nature of these links. Our results show that the digital economy has made a positive and significant contribution to sustainable development. For developing countries, mobile technologies have positive and significant effects on sustainable development, whereas, in developed countries, these effects are less elastic. The Internet has a positive and significant effect on the Human Development Index. While trade openness plays a driving role in environmental performance, it has a negative impact on equitable and sustainable development.

Keywords: Digital economy, Socio-economic, Environmental performance, Developed-Developing Countries

Introduction

The digital economy is primarily based on the use of application software, which refers to computer programs or a combination of several programs designed to perform scientific tasks, such as writing, calculation, or other operations. Computer components have enabled the

https://doi.org/10.18080/jtde.v12n3.930

execution of specific and formalised tasks (Bouri, 2019). These digital applications have been regarded as communication tools, as they have allowed brands to achieve general communication objectives. Digital marketing has provided these brands with the opportunity to engage with a large number of Internet users, thereby enhancing the digital environment. Moreover, these digital applications have had a positive impact on the business performance of companies, as products and brands attract a large number of users (Green & Sand, 2015).

In the age of global digitalisation, the pursuit of sustainable and environmentally responsible economic development has emerged as a key driver of economic growth across various countries. The digital economy has transformed traditional frameworks and methodologies for resource allocation, reducing the distortions caused by information asymmetry and limited rationality, and addressing new challenges associated with resource management due to inadequate regulation and technological misuse. The rapid advancement of the digital economy plays a pivotal role in driving technological innovation, accelerating the modernisation of industrial structures, boosting regional economic growth, and fostering social development (Han *et al.*, 2023; Li *et al.*, 2020; Ma & Zhu, 2022; Sui & Rejeski, 2002). In the post-pandemic era, the digital economy has become an essential tool for mitigating the impacts of COVID-19 and enhancing economic stability (Irfan *et al.*, 2021). The pandemic has also intensified public demand for a shift from rapid economic expansion to a more sustainable and environmentally conscious development model (Liu *et al.*, 2022). Thus, accelerating the green transition is imperative to achieving long-term sustainable development.

The digital economy acts as a catalyst for sustainable development by fostering integration between digital and real economies. This synergy encourages the advancement of digital industrial clusters, thus forging a new paradigm of economic development and promoting sustainable progress. The profound impact of this digital economy is undeniable, exerting considerable influence on real economic growth. Indeed, it has become a significant driver of sustainable development (Jing & Sun, 2019). Furthermore, the digital economy plays a crucial role in the construction of industrial agglomerations and the formation of industrial zones. These industrial concentrations also have a significant impact on sustainable development (Kai *et al.*, 2023).

The impact of the digital economy on sustainable development has become a prominent topic in recent scientific research. Savastano *et al.* (2022) explored this impact from a business perspective, while Cook *et al.* (2022) concentrated on digital agriculture as a means to support sustainable food systems. Several studies have also examined digitalisation within the framework of the circular economy, such as the works by Moreno *et al.* (2019), Denicolai *et al.* (2021), Kusi-Sarpong *et al.* (2021), and Chauhan *et al.* (2022), which are integral to sustainable development. In essence, the digital economy has infiltrated nearly every facet of life, including matters concerning sustainable development. Consequently, it is crucial to underscore the connection between the digital economy and sustainable development.

The global population is now approaching eight billion inhabitants, exacerbating the contradiction between development and the environment (Hickel, 2019). With the continued growth of the population, issues such as air and marine pollution (Shao *et al.*, 2020), climate change, resource depletion, and biodiversity loss have emerged, persistently hindering the progress of human society and impeding the path towards sustainable development. Environmental protection is an essential element of sustainable policy leading to sustainable development, thus promoting the harmonious progress of human society in symbiosis with nature (Ruggerio, 2021).

Furthermore, the digital economy contributes to promoting environmental sustainability through three specific points. Firstly, it offers the prospect of green production by leveraging technological innovations, such as virtual reality, databases, the Internet, and the digital economy (Viturka *et al.*, 2013; Jiang *et al.*, 2022; Guo *et al.*, 2023). Secondly, traditional regulatory models are facing shortcomings in regulation, out-dated measures, and low efficiency. The digital economy provides both a solid opportunity and technical advantages that lead to the adoption of higher ecological standards (Liu *et al.*, 2017; Chen *et al.*, 2018; Liu *et al.*, 2021; Li & Yang, 2024). Thirdly, the perspective of social-environmental monitoring: the public, characterised by its widespread distribution, influence power, and rapid awareness, can play a pivotal role in environmental monitoring (Dzwigol *et al.*, 2023). The convergence of these three points leads to the conclusion that the digital economy promotes both economic and environmental development.

In our study, we seek to analyse the impact of the digital economy on sustainable development through a comparative approach between developed and developing countries. Our research question focuses on the following: to what extent does the digital economy influence sustainable development? The central objective of our paper is to highlight the contribution of the digital economy to socio-economic development and its influence on environmental preservation. We also examine the impact of this digital economy on sustainable development in the context of current geopolitical risks.

This paper is organised in two parts. In the first part, we present a theoretical context and review the literature regarding the relationship between the Human Development Index (HDI), the Environmental Performance Index (EPI), and the digital economy, in order to explore how one index influences the other. In the second part, we empirically validate the static relationships that describe the impact of the digital economy on sustainable

https://doi.org/10.18080/jtde.v12n3.930

development for developed and developing countries, and we conclude this study with the findings and their implications for economic decision-makers.

Literature Review

Sustainable development results from a dynamic interaction among economic, social, educational, and environmental activities (Ucal & Xydis, 2020), leading to improved supply efficiency, achieving equitable development, creating an ecological civilisation, human modernisation, and other aspects (Ren, 2018). Social and economic issues, such as environmental and social equity, have drawn attention to sustainable development (Hopwood *et al.*, 2005). Achieving sustainable development can be addressed from two perspectives: mitigating environmental problems; and enhancing social equity. Regarding the mitigation of environmental problems, improving energy efficiency can partially alleviate the negative impacts of environmental emissions on sustainable development, thereby promoting the latter (Rosen & Dincer, 1999). Similarly, structural industrial modernisation serves as an important driver of sustainable development (Wang & Wang, 2021). As for strengthening social equity, the emergence of block-chain technology offers convenience to various economic entities by enabling real-time, automated, and publicly verifiable transaction dematerialisation, thereby enhancing social equity (Gunay *et al.*, 2023; Ma *et al.*, 2024) and fostering sustained sustainable development.

Sustainable development has become a widely used term in development discourse, associated with various definitions, meanings, and interpretations. Taken literally, sustainable development would simply mean development that can continue indefinitely or for a given period (Dernbach, 2003; Lélé, 1991; Stoddart, 2011). Structurally, the concept can be seen as a compound expression of two words: "sustainable" and "development". Sustainable development can be defined as development that meets the needs of the present generation without compromising the ability of future generations to meet their own needs (Stoddart, 2011).

We synthesise the main previous studies on the impact of the digital economy on sustainable development. We also discuss indicators of development, access, use and implementation of the digital economy, and how these indicators can be used to assess the influence of digital factors on human development and environmental performance. Many studies underline the considerable importance of advanced technologies and the development of information systems on social development.

Sharma & Gani (2004) examined how the development gap between developed and developing countries could be attributed to the progress of information and communication technologies (ICT), which stimulate and promote socio-economic development. It is important to first focus on the influence of technological advances on the three dimensions of the HDI: health, education, and the economy. Social progress is the ability of a society to meet the basic human needs of its citizens and create the conditions for all individuals to fully realise their potential (Pérez-Castro *et al.*, 2021).

As Castillo & Dresdner (2013) point out, technological evolution allows for increased efficiency of production means, which leads to an increase in production and, consequently, to profit maximisation. This in turn allows reducing costs and offering mass access to services and products. The direct consequence is a substantial improvement in the wellbeing and quality of life of a population (Foladori, 2005). Similarly, Mokyr (1990) argued that technological progress is inseparable from economic progress. In fact, he defined technology as any technical change in the production process that increases efficiency by using fewer resources or creating better and more innovative products. Porter (1985) found that technology is embedded in the production process and that any technological change directly affects it. Given that progress on social issues does not necessarily follow economic development, other means can be used to measure wellbeing and human development.

The study by Nosiru & Sodique (2019) found a significant link between ICT technologies, human development, and poverty. ICT has proven to be very crucial for improving human development and reducing poverty in Africa. These authors also found that advances in ICT should focus on improving education and rural entrepreneurship. In addition, the study by De La Hoz-Rosales *et al.* (2019) can be used as a tool by policymakers, especially in developing countries, to support the use and implementation of ICT, at the level of individual "well-being", commercial "exports-imports", and government "economic growth". The capability approach allows us to understand how an activity or human resource can improve the quality of life. ICT is a resource that can enhance the capabilities of individuals, allowing them to live the life they want.

A recent study by Pérez-Castro *et al.* (2021) analysed and compared the Human Development Index (HDI) and the Information and Communication Technologies Development Index (IDICT) of various geographical areas on a random basis. These authors sought to analyse the economic and digital divide between countries that make up a region with a very long economic and cultural history. To do this, they chose the countries bordering the Mediterranean Sea, as well as the countries of North Africa and the Middle East that have signed the recent Euro-Mediterranean association agreements. The descriptive analysis of the behaviour of the HDI and the IDICT shows that human and technological development has increased at average growth rates of 23% and 2.5%, respectively. Imran *et al.* (2022) used a panel data analysis on annual data for a sample of 28 European Union countries over the period 2018 to 2021 to study the impact of the Digital Economy and Society Index (DESI) on the Sustainable Development Goals Index (SDGI). The variables used in the study include indicators such as: the availability of Internet services, participation in online courses, the use of online banking, the use of pre-filled forms, access to digital public services for businesses, fixed broadband coverage, very high-speed fixed broadband coverage, 4G coverage, and mobile broadband adoption. The authors concluded that connectivity, human capital, and the use of Internet services have a positive and significant impact on sustainable development growth indicators. The integration of digital technology and digital public services has a lesser but positive impact.

Zhang *et al.* (2022) used panel data from 31 countries over the years 2009 to 2019 to show that the digital economy is an effective measure for mitigating the negative economic impact of the COVID-19 epidemic. A Global Trade Analysis Project (GTAP) model was used to examine the impact of COVID-19 on the digital industries and trade structure of countries along the "Belt and Road". The results show that the digital economy is having a positive and significant effect on the economic growth of these countries, even if the development of the digital economy is unevenly distributed between regions. Verbivska *et al.* (2023) concluded that digitisation is a priority for Ukraine and is essential for the country's economic development. The authors propose a series of measures to support this process and ensure its success. The study found that the Ukrainian economy is undergoing a full-scale digital transformation, even in the context of the ongoing war. Ukraine's participation in the EU's "Digital Europe" program is an important asset. However, the study also identified economic, institutional, and infrastructural obstacles that must be overcome to succeed in digitisation.

Vinod & Sharma (2021) found that the COVID-19 pandemic had a positive and significant impact on sustainable development, exacerbating existing challenges and creating new ones. The pandemic has had a negative impact on sustainable development, particularly on goals related to poverty, hunger, health, education and gender equality. Boichinko *et al.* (2023) used a comparative method to examine the effects of the war in Ukraine on the economies and societies of European Union member states over a period from February to September 2022. They concluded that the conflict has had a negative impact on the European Union's economy, as inflation has risen in most EU countries and economic growth has slowed. The conflict has also had a negative impact on labour markets, with unemployment rising in some countries. Furthermore, the conflict is having a negative and significant impact on sustainable development, and the environmental effects of the war are particularly severe.

Hypothesis Development

The models referenced in our study are based on a neoclassical conceptual framework, in which socio-economic development and environmental performance are achieved through endogenous growth. This growth is fuelled by green innovation, propelled by the digital economy, which promotes sustainable development. Firstly, green technological innovation facilitates the mobility of production factors and enhances resource efficiency, optimising socio-economic performance and reinforcing development sustainability. Secondly, the digital economy is intrinsically aligned with environmentally friendly industries. These green innovations in the digital domain can reduce carbon dioxide emissions, thereby stimulating the growth of green economies. They can also radically transform highly polluting and energy-intensive traditional industries, thus catalysing environmental performance (Ulucak *et al.*, 2020).

This study suggests two key hypotheses regarding the interaction involving the digital economy and sustainable development. First, (**H1**) we assume that the digital economy promotes sustainable development by driving green innovation. Second, (**H2**) we propose that the relationship amongst the digital economy and sustainable development is nonlinear, demonstrating that the impact of digital improvements on sustainability may vary in complexity and intensity depending on several factors (Xie *et al.*, 2020; Zhang *et al.*, 2023). These hypotheses aim to explore mutually the direct and nuanced effects of digital advances on sustainable development outcomes.

We sought the expertise of the World Bank, the United Nations and the International Monetary Fund to classify the 53 countries mentioned in our article into two categories: 25 developing countries and 28 developed countries. Additionally, we selected the Human Development Index (HDI) as an indicator of socio-economic development, serving as a criterion for classifying countries into developed and developing categories. Furthermore, the Environmental Performance Index (EPI) was chosen as an indicator to assess each country's efforts in combating pollution. In our reference models, we incorporated both HDI and EPI as two endogenous variables.

We have chosen multiple explanatory variables to account for the socio-economic development indicators of each country, including trade openness (TO), population growth rate (PGR), inflation rate (CPI), investment volumes through gross fixed capital formation (GFFC), as well as public expenditures in technologies (GC). Additionally, we have approximated the digital economy using a set of explanatory variables, namely mobile subscriptions (MS), Internet usage (IU), fourth and fifth generation wireless mobile technologies (4G5GWMT), high-technology exports (HTE), and exports of information and communication technology goods (ICTE).

We consulted the works of Khan *et al.* (2019), Zheng & Wang (2022), Phillips (2023) and Hwang (2023) to select the socio-economic and digital variables, as well as the endogenous

https://doi.org/10.18080/jtde.v12n3.930

and explanatory variables used in our reference models. We adopted the panel data methodology to specify the Human Development Index or the Environmental Performance Index based on the macroeconomic variables and the digital economy mentioned above for the period from 1990 to 2022 for the sample of 28 developed countries and the group of 25 developing countries. The definitions of each endogenous and explanatory variable, as well as the sources of each variable, are presented in Table 1.

Table 1. Definitions and Sources of Variables

Variable	Abbrev-iation	Definitions	Sources
Human Develop-ment Index	HDI	The United Nations Development Programme (UNDP) creates the Human Development Index (HDI), a composite index that measures each nation's degree of development, every year. The HDI is concerned with a nation's citizens' quality of life, as opposed to only economic indices. It is predicated on three key elements: 1. Life expectancy at birth: This factor takes into account a person's future circumstances, such as their ability to obtain clean water, sufficient nourishment, suitable housing, and healthcare facilities. 2. The degree of education: This includes both the availability and calibre of education, and it is vital to an individual's social and professional autonomy. 3. Gross national income per capita: This measure shows the average level of living and the population's access to culture, commerce, transportation, and other amenities. The HDI scale goes from zero to one, with one representing the highest level of development. Since per capita income was thought to be insufficient for assessing overall development, the HDI was created in 1990 to replace it. The basis for giving the HDI priority is the idea that personal development is a requirement for individual freedom. Four additional indices have been created in order to improve comprehension of the degree of development: • Women's and men's HDIs are compared using the Gender Development Index (GDI). • The Gender Inequality Index (GII) focuses on the empowerment of women. • The level of disparities is taken into consideration in the calculation of the Inequality-adjusted Human Development Index (IHDI). • The Multidimensional Poverty Index (MPI) measures various aspects of poverty, excluding the sole criterion of income.	World Bank and UNDP are the sources of HDI data.
Environ-mental Perform-ance Index	EPI	A measure of how close or far a country is from meeting set environmental policy targets is provided by the Environmental Performance Index (EPI), which rates each country's environmental performance on a scale of 0 to 100, with the worst performance for each indicator (Elliott & Esty (2023), elaborated in Yale Center for Environmental Law and Policy, YCELP). Consequently, according to Wolf *et al.* (2022) and the Yale Center for Environmental Law and Policy (YCELP) (Elliott & Esty, 2023), the EPI offers a tool for problem identification, target-setting, trend tracking, understanding outcomes, and identifying best practices.	United Nations Development Programme (UNDP) and DATA.NASA.GOV
Trade openness	TO	Trade refers to the total value of a country's exports and imports of goods and services, expressed as a percentage of its gross domestic product (GDP).	World Bank's Development Indicators (WDI)

https://doi.org/10.18080/jtde.v12n3.930

Variable	Abbrev-iation	Definitions	Sources
Population growth rate	PGR	Annual population growth rate for year *t* is the exponential rate of growth of midyear population from year *t-1* to *t*, expressed as a percentage.	World Bank's Development Indicators (WDI)
Consumer Price Index	CPI	The annual percentage change in the average consumer's cost of purchasing a basket of goods and services—which may be constant or vary at predetermined intervals, like annually—is what the consumer price index uses to calculate inflation. Typically, the Laspeyres formula is applied.	World Bank's Development Indicators (WDI)
Gross Formation of Fixed Capital	GFFC	Previously known as gross domestic investment, it involves building roads, machines, and improving land.	World Bank's Development Indicators (WDI)
Govern-ment consump-tion	GC	Includes all government current expenditures for purchases of goods and services	World Bank's Development Indicators (WDI)
Mobile Subscrip-tion	MS	Mobile-cellular telephone subscriptions per 100 inhabitants.	World Bank's Development Indicators (WDI)
Internet User	IU	Percentage of individuals using the Internet.	World Bank's Development Indicators (WDI)
Fourth and Fifth Generation of Wireless Mobile Technol-ogies	4G5GW MT	Fixed-broadband subscriptions per 100 inhabitants	World Bank's Development Indicators (WDI)
High Technology Export	THE	High-technology exports are products with high R&D intensity, such as aerospace, computers, pharmaceuticals, scientific instruments, and electrical machinery. Data are in current US dollars.	World Bank's Development Indicators (WDI)
ICT Exports	ICTE	Information and communication technology goods exports include computers and peripheral equipment, communication equipment, consumer electronic equipment, electronic components, and other information and technology goods (miscellaneous).	World Bank's Development Indicators (WDI)

Methodology

We refer to the models below to estimate the effect of the digital economy on sustainable development in developed and developing countries:

$$Log(HDI)_{it} = \alpha_i + \beta_1 Log(TO)_{it} + \beta_2 Log(PGR)_{it} + \beta_3 Log(CPI)_{it} + \beta_4 Log(GFFC)_{it} + \beta_5 Log(GC)_{it} + \beta_6 Log(MS)_{it} + \beta_7 Log(IU)_{it} + \beta_8 Log(4G5GWMT)_{it} + \beta_9 Log(HTE)_{it} + \beta_{10} Log(ICTE)_{it} + \varepsilon_{it} \quad (1)$$

$$Log(EPI)_{it} = \delta_i + \gamma_1 Log(TO)_{it} + \gamma_2 Log(PGR)_{it} + \gamma_3 Log(CPI)_{it} + \gamma_4 Log(GFFC)_{it} + \gamma_5 Log(GC)_{it} + \gamma_6 Log(MS)_{it} + \gamma_7 Log(IU)_{it} + \gamma_8 Log(4G5GWMT)_{it} + \gamma_9 Log(HTE)_{it} + \gamma_{10} Log(ICTE)_{it} + \varepsilon_{it} \quad (2)$$

We based our methodology on that of Zhao *et al.* (2020). We used five indicators to approximate the degree of digital economy of each country in each sample. Mobile subscriptions (MS) are at the heart of this economic advancement, with key factors including the prevalence or use of the Internet (IU), the fourth and fifth generation of wireless mobile technologies (4G5GWMT), high-tech exports (HTE), and ICT exports (ICTE). An increase in Internet penetration means better access for people to the Internet, allowing them to benefit from the advantages of digital information technologies and reducing the digital divide (Tian *et al.*, 2024).

Sustainable development involves meeting the needs of the current generation without compromising the ability of future generations to meet their own needs (Tian *et al.*, 2024). This requires a balance between efficient resource use, environmental protection, and long-term economic growth. To achieve economic sustainability within the framework of sustainable development, we referred to the study by Tian *et al.* (2024) to use two fundamental indicators of human development: the Human Development Index (HDI) and the Environmental Performance Index (EPI).

We studied the impact of the digital economy on the economic growth of each country in the two samples mentioned in our study, namely developed countries and developing countries. To do this, we used five macroeconomic indicators: trade openness (TO), the annual population growth rate (PGR), the inflation rate (CPI), gross fixed capital formation (GFCF) as a measure of national investment, and public spending (GC) allocated to the digital economy.

We used the panel-data methodology to investigate the effects of the digital economy on sustainable development in 28 developed countries, including Austria, Belgium, Canada, Chile, China, Cyprus, Denmark, Finland, France, Germany, Greece, Hungary, Ireland, Iceland, Israel, Italy, Japan, South Africa, Luxembourg, Norway, New Zealand, Poland, Portugal, Romania, Spain, Switzerland, United Kingdom and United States. Additionally, 25 developing countries were included: Bolivia, Brazil, Colombia, Ecuador, Egypt, El Salvador, Georgia, India, Indonesia, Jamaica, Jordan, Kazakhstan, Malaysia, Morocco, Mexico, Moldova, Paraguay, Philippines, Peru, Sri Lanka, Senegal, Thailand, Tunisia, Turkey and Zimbabwe.

We applied the natural logarithm to each explanatory and endogenous variable to approximate the data in our database and to linearise our reference models. This approach was adopted due to the non-linearity of the relationships between sustainable development, as mentioned in the studies by Xie *et al.* (2020) and Zhang *et al.* (2023), which constitutes the second hypothesis of our article.

https://doi.org/10.18080/jtde.v12n3.930

Results

The different statistical indicators for developed and developing countries are shown in Table 2 and Table 3.

Table 2. Descriptive statistics of developed countries

Variable	Observ-ation	Mean	Median	SD*	CV*	Kurtosis	Skewness
$Log(EPI)_{it}$	924	4.1305	4.1319	0.1814	0.0439	18.1782	-1.6760
$Log(HDI)_{it}$	924	3.6303	0.8690	4.8795	1.4643	317.3131	17.7129
$Log(TO)_{it}$	924	4.2724	4.2186	0.5396	0.1263	4.0529	0.3485
$Log(PGR)_{it}$	924	0.6674	0.5407	0.7405	1.1094	6.5827	0.8954
$Log(CPI)_{it}$	924	5.2647	2.1833	2.9024	4.7335	310.0368	15.7447
$Log(GFFC)_{it}$	924	3.1135	3.1043	0.2071	0.0665	6.0217	0.7388
$Log(GC)_{it}$	924	2.9107	2.9423	0.2181	0.0749	3.0723	-0.6419
$Log(MS)_{it}$	924	3.5531	4.5422	2.0318	0.5718	7.8727	-2.1017
$Log(IU)_{it}$	924	2.8244	4.0222	2.5709	0.9103	10.0085	-1.2199
$Log(4G5GWMT)$	924	0.7120	2.7289	3.5510	1.9872	3.5325	-1.1487
$Log(HTE)_{it}$	924	2.5681	2.7290	0.7548	0.2939	4.1031	-0.9459
$Log(ICTE)_{it}$	924	1.0684	5.5569	0.3214	0.3642	920.0011	30.3150

* (SD) Standard deviation and (CV) Coefficient of variation

Statistical indicators for developed countries show that means are low for the various macroeconomic variables and the digital economy. Moreover, standard deviations are minimal for these different variables, and their linear fits are improved. Information is asymmetrical for each of these variables, as the Skewness statistic is strictly non-zero. Non-normality is a dominant feature for these variables.

Table 3. Descriptive statistics of developing countries

Variable	Observ-ation	Mean	Median	SD*	CV*	Kurtosis	Skewness
$Log(EPI)_{it}$	825	3.8907	3.9160	0.2181	0.0561	4.8953	1.4351
$Log(HDI)_{it}$	825	2.6789	0.6790	33.8091	1.6203	2.8045	0.3310
$Log(TO)_{it}$	825	4.1779	4.1858	0.4757	0.1139	3.1803	2.7188
$Log(PGR)_{it}$	825	1.3254	1.3728	1.1024	0.8318	1.8059	-9.0806
$Log(CPI)_{it}$	825	3.2686	5.0307	1.9211	0.6554	3.6783	-2.4310
$Log(GFFC)_{it}$	825	3.1135	3.0993	0.3147	0.1026	1.9802	0.6934
$Log(GC)_{it}$	825	2.5768	2.5811	0.3060	0.1187	7.4318	0.4029
$Log(MS)_{it}$	825	2.1602	3.8408	3.3616	1.5561	3.8850	-10.5458
$Log(IU)_{it}$	825	0.8730	2.3513	3.6565	4.1882	3.2314	-10.4693
$Log(4G5GWMT)_{it}$	825	2.0065	-0.6433	4.1628	-2.0746	2.1987	-15.2852
$Log(HTE)_{it}$	825	1.6858	1.7527	1.5497	0.9192	2.6500	-3.5066
$Log(ICTE)_{it}$	825	3.1494	4.6378	0.1547	2.7227	2.0012	0.0000

* (SD) Standard deviation and (CV) Coefficient of variation

Descriptive statistics for developing countries for the various macroeconomic and numerical variables indicate non-normality, since the information is asymmetrical and not flattened. However, the linear fit of each variable, except HDI, is improved because the standard deviation is very low. Means are also very small and very close to medians, except for 4G5GWMT. Although the focus is on the overall impact of digital economy indicators on environmental performance and human development, we have also analysed this link using different sub-categories of developed and developing countries classified according to income.

https://doi.org/10.18080/jtde.v12n3.930

In Tables 2 and 3, we noted significant differences between the two groups. Developed countries have higher EPI and HDI indicators and digital economy indicators twice as high as developing countries.

The dependency relationships between the two sustainable development indicators and the digital economy variables have been studied based on the unconditional correlation matrices presented Table 4 and Table 5.

Table 4. Unconditional Correlation Matrix for Developed Countries

Variable	LEPI	LHDI	LTO	LPGR	LCPI	LGFFC
LEPI	1.0000	-	-	-	-	-
LHDI	0.1637*	1.0000	-	-	-	-
LTO	**0.3157***	**0.0466***	1.0000	-	-	-
LPGR	-0.0781**	**0.0215*****	0.2247*	1.0000	-	-
LCPI	-0.3632*	-0.0642***	-0.2228*	0.0816**	1.0000	-
LGFFC	-0.3460*	-0.0941***	-0.1544*	0.0535**	0.0583**	1.0000
LGC	**0.2099***	**0.0271*****	0.0705*	-0.1422*	-0.1079*	-0.3527*
LMS	0.3561*	0.1370**	0.3126*	0.0454**	-0.5187*	-0.0419**
LIU	0.3951*	0.1480**	0.2882*	0.0465**	-0.4789*	-0.0859**
L4G5GWMT	0.4027*	0.1441**	0.3465*	0.0898**	-0.4378*	-0.0197**
LHTE	0.0636***	0.0184***	-0.0948**	0.0284**	-0.3188*	0.1918*
LICTE	0.1911*	0.1096**	0.2948*	0.1742*	-0.3405*	-0.0723*

Variable	LGC	LMS	LIU	L4G5GWMT	LHTE	LICTE
LEPI	-	-	-	-	-	-
LHDI	-	-	-	-	-	-
LTO	-	-	-	-	-	-
LPGR	-	-	-	-	-	-
LCPI	-	-	-	-	-	-
LGFFC	-	-	-	-	-	-
LGC	1.0000	-	-	-	-	-
LMS	0.2074*	1.0000	-	-	-	-
LIU	0.2162*	0.9261*	1.0000	-	-	-
L4G5GWMT	0.2416*	0.8474*	0.8352*	1.0000		-
LHTE	0.1006*	0.4139*	0.4031*	0.3306*	1.0000	-
LICTE	0.2331*	0.5317*	0.5267*	0.5054*	0.3169*	1.0000

(*) significance at 1% risk; (**) significance at 5%; (***) significance at 10%.

For developed countries, the environmental performance index is positively correlated with trade openness and public consumption; on the other hand, this index is negatively correlated with population growth, inflation, and investment. For the human development index, we note that this index is positively correlated with the population; on the other hand, it is negatively correlated with the other macroeconomic control variables, which are trade openness, inflation, investment, and public consumption. Moreover, we can clearly see that the environmental performance index is positively correlated with the digital economy variables, such as mobile subscription, Internet users, broadband subscriptions, and high technology exports, except for the ICT export variable. Furthermore, the human development index is positively correlated with these variables of the digital economy except the last two variables, namely high technology exports and ICT exports, which are negatively correlated with the human development index. In addition, we found that there is a low correlation between macroeconomic and digital economy variables.

Table 5. Unconditional Correlation Matrix for Developing Countries

Variable	LEPI	LHDI	LTO	LPGR	LCPI	LGFFC
LEPI	1.0000	-	-	-	-	-
LHDI	0.0571***	1.0000	-	-	-	-
LTO	0.0539***	0.0820**	1.0000	-	-	-
LPGR	-0,0726***	-0,1035**	-0,1232*	1.0000	-	-
LCPI	-0,0943***	0,0273**	-0,2547*	0,0820**	1.0000	-
LGFFC	-0,1488**	0,0536**	0,1232*	0,0862***	0,0623	1.0000
LGC	0,0178***	0,0356**	0,0606**	-0,0917**	-0,2365	-0,0118***
LMS	0,1753**	0,1182*	0,1370**	-0,1976*	-0,2896	0,0587***
LIU	0,1658**	0,1217*	0,1076*	-0,2066*	-0,2833	-0,0071***
L4G5GWMT	0,1819**	0,0808**	0,0780**	-0,2326*	-0,2724	-0,0033***
LHTE	0,0941***	0,0161**	0,1104**	0,0424**	-0,1528	0,1020**
LICTE	0,1159***	-0,0564**	-0,2379**	-0,1829*	0,1097	-0,0061***

Variable	LGC	LMS	LIU	L4G5GWMT	LHTE	LICTE
LEPI	-	-	-	-	-	-
LHDI	-	-	-	-	-	-
LTO	-	-	-	-	-	-
LPGR	-	-	-	-	-	-
LCPI	-	-	-	-	-	-
LGFFC	-	-	-	-	-	-
LGC	1.0000	-	-	-	-	-
LMS	0,0789***	1.0000	-	-	-	-
LIU	0,1313**	0,9565*	1.0000	-	-	-
L4G5GWMT	0,1201**	0,8697*	0,8935*	1.0000	-	-
LHTE	-0,0800***	0,3365*	0,3119*	0,2150*	1.0000	-
LICTE	-0,2888*	0,3191*	0,3225*	0,3051*	0,1190*	1.0000

(*) significance at 1% risk; (**) significance at 5%; (***) significance at 10%

As for developing countries, the environmental performance index is positively correlated with trade openness and purchasing power; on the other hand, this index is negatively correlated with population, inflation and investment. Additionally, we show that the human development index is positively correlated with trade openness, inflation, investment and public consumption, except for population growth. Interestingly, we note that the environmental performance index is positively correlated with all variables of the digital economy. Furthermore, the human development index is positively correlated with the following variables of the digital economy: mobile subscriptions, Internet users, broadband subscriptions, and high technology exports, except for the ICT exports, which is negatively correlated with the human development index. In addition, we found that there is a low correlation between macroeconomic and digital economy variables. The absence of a multicollinearity problem is detected by the low correlation between the explanatory variables.

In contrast, for developing countries, the environmental performance index is positively correlated with trade openness and public consumption. Conversely, this index is negatively correlated with population growth, inflation, and investment. Furthermore, we find that the human development index is positively correlated with trade openness, inflation, investment, and public consumption, except for population growth. Interestingly, we note that the environmental performance index is positively correlated with all digital economy variables. Additionally, the human development index is positively correlated with mobile subscriptions, Internet users, broadband subscriptions, and high technology exports, except for ICT exports,

https://doi.org/10.18080/jtde.v12n3.930

which is negatively correlated with the human development index. Moreover, we found a low correlation between macroeconomic and digital economy variables. The low correlation between the explanatory variables indicates no multicollinearity issue.

Discussion

We resorted to the static panel-model methodology to determine a possible relationship by studying each indicator of sustainable development according to macroeconomic variables and the digital economy. To do this, we used appropriate techniques as well as the Hausman (1978) test, which allowed us to determine the nature of the individual effects. We began by estimating a static fixed effects model for conducting the Fisher test and Hausman (1978) test. Overall, the fixed effects model can control for the country-specific and year-specific impacts, as well as alleviate the problem of multicollinearity. The first step of the econometric procedure included the testing of the appropriateness of the model. This proved to be particularly useful in the discussion of the results.

Table 6. Estimation of Long-Term Relationship

Developing countries

	Within	**GLS**	**Within**	**GLS**
	LEPI	**LEPI**	**LHDI**	**LHDI**
$Log(TO)_{it}$	**0.0424****	0.0460*	**0.0312***	0.0343**
$Log(PGR)_{it}$	**0.0019**	0.0083	**-0.0903**	-0.079**
$Log(CPI)_{it}$	**-0.0168**	-0.0046	**0.0245**	0.0314**
$Log(GFFC)_{it}$	**-0.1411*****	-0.1343***	**0.0555***	0.0797**
$Log(GC)_{it}$	**0.0402**	-0.0420*	**-0.0425***	-0.0218*
$Log(MS)_{it}$	**0.0252**	0.0138*	**0.0135***	0.0048*
$Log(IU)_{it}$	**0.0212**	-0.0104*	**0.0589***	0.0355*
$Log(4G5GWMT)_{it}$	**0.0176**	0.0104*	**-0.0339***	-0.0168*
$Log(HTE)_{it}$	**0.0044**	0.0092*	**-0.0098***	-0.0081*
$Log(ICTE)_{it}$	**0.0167**	0.0127*	**-0.0247***	-0.0256*
Stat-Hausman	49.91 (0.0000)		40.64 (0.0000)	

Developed countries

	Within	**GLS**	**Within**	**GLS**
	LEPI	**LEPI**	**LHDI**	**LHDI**
$Log(TO)_{it}$	**0.0544****	0.0448*	0.1119	**-0.0308***
$Log(PGR)_{it}$	**-0.0014***	-0.0023	-0.0111	**0.0034**
$Log(CPI)_{it}$	**-0.0288***	-0.0313*	0.0052	**0.0046**
$Log(GFFC)_{it}$	**-0.001**	-0.0706**	0.0374	**-0.2375*****
$Log(GC)_{it}$	**-0.1505*****	-0.0192	-0.1356*	**-0.1018**
$Log(MS)_{it}$	**-0.0255***	0.0056	-0.0262	**-0.0029**
$Log(IU)_{it}$	**0.0039**	-0.0056	0.0072	**0.0138***
$Log(4G5GWMT)_{it}$	**0.0200***	0.01738*	0.0183	**0.0124***
$Log(HTE)_{it}$	**-0.0227***	-0.0237*	-0.0261	**-0.0213**
$Log(ICTE)_{it}$	**0.0255***	0.01464	0.0419	**0.0224***
Stat-Hausman	109.96 (0.0000)		9.50 (0.4853)	

(*) Significance at the 1% level; (**) Significance at the 5% level; (***) Significance at the 10% level.

The Hausman (1978) test indicates (Table 6) that the two static relationships estimated by Within and Generalised Least Squares (GLS) have random individual effects, as the Hausman (1978) statistics are statistically insignificant. Therefore, we retained and interpreted the

https://doi.org/10.18080/jtde.v12n3.930

results obtained by the GLS method. We noted that the control variables, including the Consumer Price Index, public consumption, and private investment, negatively and significantly affect the degree of environmental health, improve ecosystem health, and mitigate climate change in both developed and developing countries. Similarly, we found that a 1% increase in trade openness is associated with an average increase in environmental performance by 0.04% in both groups of countries. For developed countries, trade openness had a negative coefficient, suggesting that greater trade openness is associated with lower ICT contributions to human development. On the other hand, this variable had a positive and significant impact on human development in developing countries. Our results are consistent with the findings of Tamazian & Rao (2010).

We find that higher levels of private investment help to improve the dimensions of human development: a long and healthy life; being knowledgeable; and achieving a decent standard of living. In developing countries, this contributes to equitable and sustainable development. For every 1% increase in purchasing power, there is an average increase of 0.079 in socio-economic development. On the other hand, in developed countries, private investment has a negative and significant impact on human development, with a 10% decrease associated with this variable. Inflation can be harmful if not accompanied by a strong institutional framework. In this sense, it is noteworthy that governments can help to improve the three dimensions of environmental performance by establishing robust policy and institutional structures. These structures have long-term benefits in reducing greenhouse gas emissions, leading to an improvement in environmental performance. Additionally, governments should support the development of new technologies and Internet infrastructure that contribute to a more sustainable economy. We find that every 1% increase in population is associated with an average increase of 0.008% in the environmental performance of developing countries. These findings are similar to those obtained by Jain & Nagpal (2019), who found a positive association between the EPI and human capital in the context of South Asian countries. This suggests that, as human capital accumulates, it leads to lower environmental damage and better environmental performance, contrary to the case of many developing countries (Phillips, 2023). On the other hand, in developed countries, every 1% increase in population is associated with a 0.079% decrease in human development.

Nor is the effect of population growth on human development very clear from a theoretical point of view. On the one hand, some arguments suggest that a larger population could have a positive impact on economic growth, as it increases consumption needs and facilitates economies of scale. On the other hand, others argue that a larger population has to manage a fixed number of natural resources, resulting in lower prosperity. In fact, the results presented in Table 5 suggest a strong negative correlation between human development and the rate of

https://doi.org/10.18080/jtde.v12n3.930

population growth. Developing countries are more affected (resulting in lower levels of human development) due to their higher population growth rates than developed countries. These results are consistent with the arguments presented by Weil (2013) and Myovella *et al.* (2019).

Furthermore, with regard to digitisation variables, mobile subscriptions, 4G and 5G mobile technology, high-tech exports (HTE) and ICT exports (ICTE) have a positive and significant impact on environmental performance in developing countries, with the exception of Internet users. In addition, cell phone subscriptions and Internet users show positive and significant contributions of 10% associated with socio-economic development increases of 0.0048 and 0.0355, respectively. Notably, a 1% increase in the number of Internet users is associated with an average 0.03% increase in human development. However, the coefficients for broadband subscriptions, HTE and ICTE are negative. This result confirms the findings of Imran *et al.* (2022) that connectivity, human capital and the use of Internet services have a positive impact on sustainable development growth indicators. The integration of digital technology and digital public services has a lesser but positive impact.

We have observed that broadband subscriptions (MS), HTE, and ICTE have negative impacts on the sustainable development of developed countries (HDI). Consequently, these exports of high technologies and ICT have negative repercussions on the socio-economic development index of these countries, as there is a technological drain towards developing countries that offer tax advantages. This drain leads to a reduction in technological investments, thus causing economic stagnation in developed countries. However, this interpretation contradicts the study by Tian *et al.* (2024), which showed a positive influence of these technological and ICT exports on socio-economic development in the case of China.

The development of the digital economy in developed countries promotes high-quality adjustment of industrial structure, improves manufacturing efficiency, and simultaneously saves energy and reduces pollutant emissions. Moreover, it advances industry to attract an inflow of skilled workers and investments and improves enterprises' efficiency in treating pollutants for abundant intermediate products and talent pools. Thus, the air pollution reduction effect of the digital economy is significant in developed countries. However, these countries have a comparatively mature economy with a reasonable industrial structure and superior production techniques. The digital economy would cultivate new businesses based on this advantage, enhance diversified agglomeration, expand the scale of industries, and accelerate the pace of green research and development and its commercial application. Therefore, diversified agglomeration plays a significant mediating role in the developed countries (Higon *et al.*, 2017).

The notable differences between the two groups of countries are that a majority of the population in the least developed countries (global South) depend on cell phones for their economic activities, compared to the population in the developed countries (global North).

At the same time, we observe that the contribution of the Internet to the development of quality of life, hope and education is positive in both developed and developing countries. Thus, in developed countries, the deployment of fixed Internet seems to play an important role as a fundamental infrastructure, not only to improve communication, as seems to be the case in developing countries, but also in almost all sectors in developed countries. This may illustrate the differences in the way people use technology, depending on a country's level of development. Whereas in less developed countries the Internet seems to play a basic communications role, in developed countries it supports all sectors of the economy. The more developed a country, the more dependent it is on Internet use, compared to less developed countries. Secondly, the spread of cell phone technologies in developing countries is largely independent of the fixed Internet infrastructure, meaning that the Internet infrastructure is undeveloped compared to that in developed countries.

In terms of environmental performance, the results show that the Internet's contribution is negative and insignificant in developed countries. In developed countries, the Internet has a negative and significant impact. As the number of Internet users and advances in technology rise, energy consumption also increases, which ultimately leads to a decrease in performance. However, Samimi *et al.* (2011) stated that, in the case of developing countries where environmental degradation was occurring, higher human development did not reciprocally lead to improved environmental performance. Lai & Chen (2020) and Tang *et al.* (2020) both contextualised the relationships described by stating that higher levels of education and income led to greater economic development, and thus to greater environmental awareness and action. Martens & Raza (2010) put the issue in a broader context, stating that EPI and HDI are closely linked to globalisation.

Theoretical and Practical Contributions

Among theoretical contributions of this study, we can site the "Expanded Understanding of Digital Economy and Sustainable Development". In fact, this study advances the theoretical framework linking digital technology diffusion to sustainable development by elucidating the positive correlations between digital economy variables and the Human Development Index (HDI) and Environmental Performance Index (EPI). The analysis confirms that technological progress is not only associated with improved human development outcomes but also contributes to enhanced environmental performance, thereby supporting theories that integrate technology with sustainability. We also recognise the "Identification of Digital Divide

Effects". The research highlights the theoretical implications of the digital divide between developed and developing countries. It demonstrates how this divide impacts the efficacy of the digital economy in driving sustainable development. By contrasting the rapid technological advances in developed countries with the slower pace in developing countries, the study adds depth to existing theories about the unequal distribution of technological benefits and its repercussions on sustainable growth. We identify the "Impact of Innovation on Sustainable Development". The findings offer new theoretical insights into how levels of innovation moderate the relationship between the digital economy and sustainable development. The study underscores that higher innovation levels enhance the positive impacts of the digital economy on sustainable development, suggesting a nuanced interaction between technological advancement and green innovation.

For practical contributions of this research, we can site "Policy Recommendations for Bridging the Digital Divide". In fact, the study provides actionable recommendations for policymakers, particularly in developing countries. It emphasises the need for improved communication technology infrastructure and digital-skills training to bridge the digital divide. By focusing on the expansion of ICT access and the promotion of open-source technologies, policymakers can enhance socio-economic growth and environmental sustainability. We also can identify the "Guidance for Enhancing Environmental and Human Development". Practical strategies derived from the study include investing in digital infrastructure to boost both human development and environmental performance. Governments should prioritise policies that integrate digital economy advances with environmental governance, thus improving overall quality of life and environmental health. The recommendations also stress the importance of fostering technological collaboration and innovation as a means to drive sustainable development. We recognise a "Strategic Focus on Green Innovation". The research underscores the need for increased investment in green innovation and digital transformation in traditional industries. By adopting digital technologies, businesses can achieve greater efficiency and contribute to environmental sustainability. This practical insight is crucial for developing countries aiming to modernise their economies while addressing environmental challenges. We acknowledge the "Consideration of Geopolitical and Cybersecurity Risks". The study highlights the impact of geopolitical risks and cybersecurity threats on the effectiveness of the digital economy in promoting sustainable development. Future research should integrate these factors into models to provide a more comprehensive understanding of their influence. Policymakers should be aware of these risks and develop strategies to mitigate their negative effects on both technological progress and sustainable development outcomes.

https://doi.org/10.18080/jtde.v12n3.930

These theoretical and practical contributions provide a robust foundation for future research and policy development, offering valuable insights into how digital technology can be leveraged to achieve sustainable development goals across different national contexts.

Conclusion

This article investigates the influence of digital technology diffusion on the degree of sustainable development. We analysed a sample of 25 developing and 28 developed countries. The aim of our research was to determine the relationship between the digital economy and two dimensions of sustainable development. In addition, we sought to assess the level of development of developed and developing countries, and to compare their respective rates of change. Our study of these countries enabled us to identify the relationships between sustainable development and the digital economy. It has also revealed how changes in one have impacted on the other.

In the correlation of two groups of developing and developed countries, the two sustainable development indicators, Human Development Index and Environmental Performance Index, and the behaviour of the digital economy variables showed that technological development and the two sustainable development indicators have progressed and have positive correlations with each other, respectively. We also noted that, in the majority of countries, there was a direct correlation between the indicators, as reflected in the strong linear correlation between the digital economy variables and the two sustainable development indicators.

Precisely for this reason, it makes sense to correlate the digital economy variables with the EPI and HDI. Increased technological infrastructure and access to ICT directly improve a population's quality of life. Investing in and promoting ICT and the digital economy are excellent ways to enhance a country's economic growth and reduce the socioeconomic divide. One way to achieve this objective is to utilise open-source software and the growing trend of open-access technology. This allows for the creation of content freely accessible to all interested users, regardless of income level. Without this option, the cost of software licenses would pose a significant barrier for people in developing countries with fewer resources. When high technology and artificial intelligence are readily accessible, the educational level of the population can be raised by providing online courses and training programs that are considerably more affordable than traditional educational institutions. Technological collaboration projects are also crucial, as they facilitate knowledge sharing and foster cooperation between public and private institutions. This type of research initiative drives innovation, leading to the creation of resources that benefit everyone.

Our study found that increased technological development has a positive impact on human development. However, as inflation increased more rapidly in developing countries, the digital economy experienced a rapid decline in those regions. Developed countries have higher EPI, HDI, and digital economy indicators compared to developing countries. Despite this overall trend, India remains among the top countries globally in terms of innovation indicators, such as IT service exports and the number of science and engineering graduates. This is further supported by India's strong performance in areas such as the quality of its universities and scientific publications, economy-wide investments, and exports of creative goods.

The development of the digital economy is having an effective and significant impact on the environmental performance of developed rather than developing countries, by reducing the emission rate of industrial sulphur dioxide and smoke dust, as well as the per capita emission level of these air pollutants, which in turn improves air quality in cities. The result is an overall improvement in environmental performance. Our results reflect the existence of a digital divide between developed and developing countries in terms of geographical development. Nevertheless, despite the overall improvement observed in the indicators of all countries, we observe a slight decrease in the differences. This can be explained by the fact that rates of change were higher in developing countries.

Our paper enhances the theoretical framework linking digital technology with sustainable development by demonstrating positive correlations between digital economy variables and human and environmental performance indicators (HDI and EPI). It also highlights how the digital divide between developed and developing countries affects the ability of the digital economy to drive sustainable development, adding depth to theories about the unequal distribution of technological benefits. The study provides new theoretical insights into how innovation levels influence the relationship between the digital economy and sustainable development, showing that higher innovation enhances these positive impacts. This research offers practical recommendations for policymakers, particularly in developing countries, focusing on improving ICT infrastructure and digital skills to bridge the digital divide and foster socio-economic and environmental benefits.

Our results provide some useful policy implications for improving environmental performance and human development worldwide. Essentially, policymakers in developing countries should prioritise the coordinated development of the digital economy and environmental governance so as to simultaneously improve environmental health, preserve ecosystem vitality, and contribute to climate-change mitigation. In developing countries, an urgent need exists to improve communication technology infrastructure, bridge the digital divide, implement comprehensive industrial restructuring and upgrading, and effectively reduce CO_2 pollution

https://doi.org/10.18080/jtde.v12n3.930

emissions by fostering the rapid development of the digital economy. Moreover, encouraging increased public participation in environmental oversight is crucial.

Research limitations and suggestions for future research

We have obtained robust results that support the driving role of the digital economy in sustainable development through green innovation. This suggests that the digital economy helps improve levels of green innovation, acting as a catalyst for long-term development in developed countries. Furthermore, the impact of the digital economy on stimulating sustainable development gradually increases under the moderating effect of innovation levels, and there is heterogeneity in this impact for developing countries.

We have found a positive correlation between digital economy and the Environmental Performance Index (EPI) for both developed and developing nations. However, this relationship is highly susceptible to current geopolitical risks, as the Russia-Ukraine and Hamas-Israel conflicts have reduced technology transfers between the countries in our sample.

We have demonstrated that the digital economy contributes only to a limited extent to the socio-economic development of both developed and developing countries. This is attributed to these countries' heavy reliance on manufacturing industries, which rely on non-renewable energy sources and raw materials. These sectors are particularly vulnerable to current geopolitical risks. Therefore, it becomes crucial to incorporate the geopolitical risk index as an explanatory variable in our models to enhance understanding of the impact of the digital economy on the Human Development Index or the Environmental Performance Index, taking macroeconomic indicators. Additionally, including indicators of cybercrime related to the digital economy would be relevant to studying their negative contribution to the sustainable development of both developed and developing countries.

It would be wise to include these cybercrime indicators as well as the Geopolitical Risk Index (GRI) in our future research to analyse both the positive and negative impacts of the digital economy, as well as geopolitical risks, on the sustainable development of both developed and developing countries.

References

Adeel-Farooq, R. M., Riaz, M. F., & Ali, T. (2021). Improving the environment begins at home: Revisiting the links between FDI and environment. *Energy*, *215*, 119150. https://doi.org/10.1016/j.energy.2020.119150

Ahmadova, G., Bueno García, M., Delgado-Márquez, B., & Pedauga, L. (2023). Firm-and country-specific advantages: Towards a better understanding of MNEs' environmental performance in the international arena. Organization & Environment, 36(3), 468–497.https://doi.org/10.1177/10860266221129699

Ahmadova, G., Delgado-Márquez, B. L., Pedauga, L. E., & Leyva-de la Hiz, D. I. (2022). Too good to be true: The inverted U-shaped relationship between home-country digitalization and environmental performance. *Ecological Economics*, *196,* 107393. https://doi.org/10.1016/j.ecolecon.2022.107393

Altig, D., Baker, S., Barrero, J. M., Bloom, N., Bunn, P., Chen, S., Davis, S. J., Leather, J., Meyer, B., Mihaylov, E., Mizen, P., Parker, N., Renault, T., Smietanka, P., & Thwaites, G. (2020). Economic uncertainty before and during the COVID-19 pandemic. *Journal of public economics*, *191*, 104274. https://doi.org/10.1016/j.jpubeco.2020.104274

Barro, R. J., & McCleary, R. M. (2003). Religion and economic growth across countries. *American sociological review*, *68*(5), 760–781. https://doi.org/10.1177/000312240306800505

Boichenko, K., Ashraf, R. U., Mata, M. N., & Gherghina, Ş. C. (2023). The EU's Socio-economic Development Against the Backdrop of the War in Ukraine. *Central European Economic Journal*, *10*(57), 72–89. https://doi.org/10.2478/ceej-2023-0005

Bouri, N. (2019). The impact of the digital economy in Algeria on socio-cultural development. *Tributaries*, *3*(2), 259–268. https://doi.org/10.37168/1957-003-002-012

Castillo, C., & Dresdner, J. (2013). Effort optimisation in artisanal fisheries with multiple management objectives, collective quotas and heterogeneous fleets. *Australian Journal of Agricultural and Resource Economics*, *57*(1), 104–122. https://doi.org/10.1111/j.1467-8489.2012.00609.x

Chauhan, C., Parida, V., & Dhir, A. (2022). Linking circular economy and digitalisation technologies: A systematic literature review of past achievements and future promises. *Technological Forecasting and Social Change*, *177*, 121508. https://doi.org/10.1016/j.techfore.2022.121508

Chen, Z., Kahn, M. E., Liu, Y., & Wang, Z. (2018). The consequences of spatially differentiated water pollution regulation in China. *Journal of Environmental Economics and Management*, *88*, 468–485. https://doi.org/10.1016/j.jeem.2018.01.010

Cook, S., Jackson, E. L., Fisher, M. J., Baker, D., & Diepeveen, D. (2022). Embedding digital agriculture into sustainable Australian food systems: pathways and pitfalls to value creation. *International Journal of Agricultural Sustainability*, *20*(3), 346–367. https://doi.org/10.1080/14735903.2021.1937881

Das Neves Almeida, T. A., Cruz, L., Barata, E., & García-Sánchez, I. M. (2017). Economic growth and environmental impacts: An analysis based on a composite index of environmental damage. *Ecological Indicators*, *76*, 119–130. https://doi.org/10.1016/j.ecolind.2016.12.028

De La Hoz-Rosales, B., Ballesta, J. A. C., Tamayo-Torres, I., & Buelvas-Ferreira, K. (2019). Effects of information and communication technology usage by individuals, businesses, and government on human development: An international analysis. *IEEE Access*, *7*, 129225–129243. https://doi.org/10.1109/ACCESS.2019.2939404

Denicolai, S., Zucchella, A., & Magnani, G. (2021). Internationalization, digitalization, and sustainability: Are SMEs ready? A survey on synergies and substituting effects among growth paths. *Technological Forecasting and Social Change*, *166*, 120650. https://doi.org/10.1016/j.techfore.2021.120650

Dernbach, J. C. (2003). Achieving sustainable development: The centrality and multiple facets of integrated decisionmaking. *Global Legal Studies*, *10*(1), 247–284. https://doi.org/10.2979/gls.2003.10.1.247

Dzwigol, H., Kwilinski, A., Lyulyov, O., & Pimonenko, T. (2023). The role of environmental regulations, renewable energy, and energy efficiency in finding the path to green economic growth. *Energies*, *16*(7), 3090. https://doi.org/10.3390/en16073090

Elliott, E. D., & Esty, D. C. (2023). Environmental Law for the 21st Century. *Pace Environmental Law Review*, *40*(3), 454. https://doi.org/10.58948/0738-6206.1877

Elsalih, O., Sertoglu, K., & Besim, M. (2020). Environmental performance, comparative advantage of crude oil and the role of institutional quality. *Environmental science and pollution research*, *27*, 3489–3496. https://doi.org/10.1007/s11356-019-06838-9

Foladori, G. (2005). Modernización ecológica, cambio tecnológico y globalización. *Economía, sociedad y territorio*, *18*, 335–353. Available at: http://ricaxcan.uaz.edu.mx/jspui/handle/20.500.11845/233

GeSI. (2019). Digital with Purpose: Delivering a SMARTer2030. Available at https://gesi.org/research/gesi-digital-with-purpose-full-report#:~:text=The%20report%2C%20Digital%20with%20Purpose,each%20of%20the%2017%20SDGs

Green, D. A., & Sand, B. M. (2015). Has the Canadian labour market polarized? *Canadian Journal of Economics/Revue canadienne d'économique*, *48*, 612–646. https://doi.org/10.1111/caje.12145

Gunay, S., Sraieb, M. M., Kaskaloglu, K., & Yıldız, M. E. (2023). Cryptocurrencies and global sustainability: do blockchained sectors have distinctive effects?. *Journal of Cleaner Production*, *425*, 138943. https://doi.org/10.1016/j.jclepro.2023.138943

Guo, B., Wang, Y., Zhang, H., Liang, C., Feng, Y., & Hu, F. (2023). Impact of the digital economy on high-quality urban economic development: Evidence from Chinese cities. *Economic Modelling*, *120*, 106194. https://doi.org/10.1016/j.econmod.2023.106194

Han, H., Shiwakoti, R. K., Jarvis, R., Mordi, C., & Botchie, D. (2023). Accounting and auditing with blockchain technology and artificial Intelligence: A literature review. *International Journal of Accounting Information Systems*, *48*, 100598. https://doi.org/10.1016/j.accinf.2022.100598

Hao, Y., Wu, Y., Wang, L., & Huang, J. (2018). Re-examine environmental Kuznets curve in China: Spatial estimations using environmental quality index. *Sustainable Cities and Society*, *42*, 498–511. https://doi.org/10.1016/j.scs.2018.08.014

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica*, *46*(6), 1251–1271. https://doi.org/10.2307/1913827

Hickel, J. (2019). The contradiction of the sustainable development goals: Growth versus ecology on a finite planet. *Sustainable Development*, *27*(5), 873–884. https://doi.org/10.1002/sd.1947

Higón, D. A., Gholami, R., & Shirazi, F. (2017). ICT and environmental sustainability: A global perspective. *Telematics and Informatics*, *34*(4), 85–95. https://doi.org/10.1016/j.tele.2017.01.001

Hopwood, B., Mellor, M., & O'Brien, G. (2005). Sustainable development: mapping different approaches. *Sustainable development*, *13*(1), 38–52. https://doi.org/10.1002/sd.244

Howson, P. (2021). Distributed degrowth technology: Challenges for blockchain beyond the green economy. *Ecological Economics*, *184*, 107020. https://doi.org/10.1016/j.ecolecon.2021.107020

Hwang, Y. K. (2023). The synergy effect through combination of the digital economy and transition to renewable energy on green economic growth: Empirical study of 18 Latin American and Caribbean countries. *Journal of Cleaner Production*, *418*. https://doi.org/10.1016/j.jclepro.2023.138146

Imran, M., Liu, X., Wang, R., Saud, S., Zhao, Y., & Khan, M. J. (2022). The Influence of Digital Economy and Society Index on Sustainable Development Indicators: The Case of European Union. *Sustainability*, *14*(18), 11130. https://doi.org/10.3390/su141811130

Irfan, M., Hao, Y., Ikram, M., Wu, H., Akram, R., & Rauf, A., (2021). Assessment of the public acceptance and utilization of renewable energy in Pakistan. *Sustainable Production and Consumption*, *27*, 312–324. https://doi.org/10.1016/j.spc.2020.10.031

Jain, M., & Nagpal, A. (2019). Relationship between environmental sustainability and human development index: A case of selected South Asian Nations. *Vision*, *23*(2), 125–133. https://doi.org/10.1177/0972262919840202

Jiang, Q., Rahman, Z. U., Zhang, X., & Islam, M. S. (2022). An assessment of the effect of green innovation, income, and energy use on consumption-based CO2 emissions: Empirical evidence from emerging nations BRICS. *Journal of Cleaner Production*, *365*, 132636. https://doi.org/10.1016/j.jclepro.2022.132636

Jing, W. J., & Sun, B. W. (2019). Digital economy promotes high-quality economic development: A theoretical analysis framework. *Economist*, *2*, 66–73. https://doi.org/10.16158/j.cnki.51-1312/f.2019.02.008

Khan, S. A. R., Sharif, A., Golpîra, H., & Kumar, A. (2019). A green ideology in Asian emerging economies: From environmental policy and sustainable development. *Sustainable development*, *27*(6), 1063–1075. https://doi.org/10.1002/sd.1958

Kai, M. L., Limu, S., Chang, Y., & Aishan, W. (2023). Manufacturing industry agglomeration, environmental regulation and high-quality economic development. *Statistics and Decision-making*, *39*(09), 125-130. https://doi.org/10.13546/j.cnki.tjyjc.2023.09.022

Kusi-Sarpong, S., Gupta, H., Khan, S. A., Chiappetta Jabbour, C. J., Rehman, S. T., & Kusi-Sarpong, H. (2021). Sustainable supplier selection based on industry 4.0 initiatives within the context of circular economy implementation in supply chain operations. *Production Planning & Control*, *34*(10), 999–1019. https://doi.org/10.1080/09537287.2021.1980906

Lai, S. L., & Chen, D. N. (2020). A research on the relationship between environmental sustainability management and human development. *Sustainability*, *12*(21), 9001. https://doi.org/10.3390/su12219001

Lange, S., & Santarius, T. (2020). *Smart Green World. Making Digitalization Work for Sustainability* (1st ed.). Routledge. . https://doi.org/10.4324/9781003030881

Lange, S., Pohl, J., & Santarius, T. (2020). Digitalization and energy consumption. Does ICT reduce energy demand? *Ecological economics*, *176*, 106760. https://doi.org/10.1016/j.ecolecon.2020.106760

Lélé, S. M. (1991). Sustainable development: a critical review. *World development*, *19*(6), 607–621. https://doi.org/10.1016/0305-750X(91)90197-P

Li, F. (2020). Leading digital transformation: three emerging approaches for managing the transition. *International Journal of Operations & Production Management*, *40*(6), 809–817. https://doi.org/10.1108/IJOPM-04-2020-0202

Li, H., & Yang, Z. (2024). Does digital economy development affect urban environment quality: Evidence from 285 cities in China. *Plos One*, *19*(2), e0297503. https://doi.org/10.1371/journal.pone.0297503

Lim, W. M., Chin, M. W. C., Ee, Y. S., Fung, C. Y., Giang, C. S., Heng, K. S., Kong, M. L. F., Lim, A. S. S., Lim, B. C. Y., Lim, R. T. H., Lim T. Y., Ling C. C., Mandrinos, S., Nwobodo, S., Phany, C. S. C., She, L., Sim, C. H., Su, S. I., Wee, G. W. E., & Weissmann, M. A. (2022). What is at stake in a war? A prospective evaluation of the Ukraine and Russia conflict for business and society. *Global Business and Organizational Excellence*, *41*(6), 23–36. https://doi.org/10.1002/joe.22162

Liu, Y., Peng, Y., Wang, B., Yao, S., & Liu, Z. (2017). Review on cyber-physical systems. *IEEE/CAA Journal of Automatica Sinica*, *4*(1), 27–40. https://doi.org/10.1109/JAS.2017.7510349

Liu, Z., Liu, J., & Osmani, M. (2021). Integration of digital economy and circular economy: Current status and future directions. *Sustainability*, *13*(13), 7217. https://doi.org/10.3390/su13137217

Liu, Y., Yang, Y., Li, H., & Zhong, K. (2022). Digital Economy Development, Industrial Structure Upgrading and Green Total Factor Productivity: Empirical Evidence from China's Cities. *International Journal of Environmental Research and Public Health*, *19*(4), 2414. https://doi.org/10.3390/ijerph19042414

Ma, X., Feng, X., Fu, D., Tong, J., & Ji, M. (2024). How does the digital economy impact sustainable development?—An empirical study from China. *Journal of Cleaner Production*, *434*, 140079. https://doi.org/10.1016/j.jclepro.2023.140079

Ma, D., & Zhu, Q. (2022). Innovation in emerging economies: Research on the digital economy driving high-quality green development. *Journal of Business Research*, *145*, 801–813. https://doi.org/10.1016/j.jbusres.2022.03.041

Martens, P., & Raza, M. (2010). Is globalisation sustainable? *Sustainability*, *2*(1), 280–293. https://doi.org/10.3390/su2010280

Mavragani, A., Nikolaou, I. E., & Tsagarakis, K. P. (2016). Open economy, institutional quality, and environmental performance: A macroeconomic approach. *Sustainability*, *8*(7), 601. https://doi.org/10.3390/su8070601

Mokyr, J. (1990). Punctuated equilibria and technological progress. *The American Economic Review*, *80*(2), 350–354. https://www.jstor.org/stable/2006599

Moreno, M., Court, R., Wright, M., & Charnley, F. (2019). Opportunities for redistributed manufacturing and digital intelligence as enablers of a circular economy. *International Journal of Sustainable Engineering*, *12*(2), 77–94. https://doi.org/10.1080/19397038.2018.1508316

Myovella, G., Karacuka, M., & Haucap, J. (2020). Digitalization and economic growth: A comparative analysis of Sub-Saharan Africa and OECD economies. *Telecommunications Policy*, *44*(2), 101856. https://doi.org/10.1016/j.telpol.2019.101856

Nasierowski, W., & Arcelus, F. J. (2003). On the efficiency of national innovation systems. *Socio-Economic Planning Sciences*, *37*(3), 215–234. https://doi.org/10.1016/S0038-0121(02)00046-0

Nosiru, O. M., & Sodique, F. R. (2019). ICT, development and poverty nexus in Africa: way forward. *Ethiopian e-Journal for Research and Innovation Foresight (Ee-JRIF)*, *11*(1). Available at: https://journals.bdu.edu.et/index.php/eejrif4/article/view/328

Pérez-Castro, M. Á., Mohamed-Maslouhi, M., & Montero-Alonso, M. Á. (2021). The digital divide and its impact on the development of Mediterranean countries. *Technology in Society*, *64*, 101452. https://doi.org/10.1016/j.techsoc.2020.101452

Phillips, J. (2023). Determining sustainability using the Environmental Performance Index and Human Development Index–An alternative approach to the Environmental Human Index through a holistic quantitative dynamic framework. *Science of The Total Environment*, *884*, 163752. https://doi.org/10.1016/j.scitotenv.2023.163752

Porter, M. E. (1985). Technology and competitive advantage. *Journal of business strategy*, *5*(3), 60–78. https://doi.org/10.1108/eb039075

Ren, B. P. (2018). China's economy changes from high-speed growth to high-quality development in the new era: Theoretical interpretation and practical orientation. *Academic monthly*, *3*, 66–86. Available at: https://www.xsyk021.com/article/id/4afc4eef-a41d-4cf1-9a92-5ab9974db448

Rosen, M. A., & Dincer, I. (1999). Exergy analysis of waste emissions. *International Journal of Energy Research*, *23*(13), 1153–1163. https://doi.org/10.1002/(SICI)1099-114X(19991025)23:13%3C1153::AID-ER545%3E3.0.CO;2-Y

Ruggerio, C. A. (2021). Sustainability and sustainable development: A review of principles and definitions. *Science of the Total Environment*, *786*, 147481. https://doi.org/10.1016/j.scitotenv.2021.147481

Samimi, A. J., Sadeghi, S., & Sadeghi, S. (2011). Tourism and economic growth in developing countries: P-VAR approach. *Middle-East journal of scientific research*, *10*(1), 28–32. Available at: https://www.cabidigitallibrary.org/doi/full/10.5555/20123020386

Savastano, M., Zentner, H., Spremić, M., & Cucari, N. (2022). Assessing the relationship between digital transformation and sustainable business excellence in a turbulent scenario. *Total Quality Management & Business Excellence*, 1–22. https://doi.org/10.1080/14783363.2022.2063717

Shao, Z., Ding, L., Li, D., Altan, O., Huq, M. E., & Li, C. (2020). Exploring the relationship between urbanization and ecological environment using remote sensing images and statistical data: A case study in the Yangtze River Delta, China. *Sustainability*, *12*(14), 5620. https://doi.org/10.3390/su12145620

Sharma, B., & Gani, A. (2004). The effects of foreign direct investment on human development. *Global economy journal*, *4*(2), 1850025. https://doi.org/10.2202/1524-5861.1049

Stoddart, H. (2011). *A Pocket guide to sustainable development governance*. Stakeholder Forum, Commonwealth Secretariat Available at: http://www.blogdocancado.com/wp-content/uploads/2011/02/Guia-de-bolso-para-a-governanca-voltada-para-o-desenvolvimento-sustentavel-ingles.pdf

Sui, D., & Rejeski, D. (2002). Environmental Impacts of the Emerging Digital Economy: The E-for-Environment E-Commerce? *Environmental Management*, *29*, 155–163. https://doi.org/10.1007/s00267-001-0027-X

Tamazian, A., & Rao, B. B. (2010). Do economic, financial and institutional developments matter for environmental degradation? Evidence from transitional economies. *Energy economics*, *32*(1), 137–145. https://doi.org/10.1016/j.eneco.2009.04.004

Tang, B., Wang, X., Li, Q., Bragazzi, N. L., Tang, S., Xiao, Y., & Wu, J. (2020). Estimation of the transmission risk of the 2019-nCoV and its implication for public health interventions. *Journal of clinical medicine*, *9*(2), 462. https://doi.org/10.3390/jcm9020462

Tian, H., Qin, J., Cheng, C., Javeed, S. A., & Chu, T. (2024). Towards low-carbon sustainable development under Industry 4.0: The influence of industrial intelligence on China's carbon mitigation. *Sustainable Development*, *32*(1), 455–480. https://doi.org/10.1002/sd.2664

Ucal, M., & Xydis, G. (2020). Multidirectional relationship between energy resources, climate changes and sustainable development: Technoeconomic analysis. *Sustainable Cities and Society*, *60*, 102210. https://doi.org/10.1016/j.scs.2020.102210

Ulucak, R., Danish, & Khan, S. U. D. (2020). Does information and communication technology affect CO2 mitigation under the pathway of sustainable development during the mode of globalization? *Sustainable Development*, *28*(4), 857–867. https://doi.org/10.1002/sd.2041

Verbivska, L., Abramova, M., Gudz, M., Lyfar, V., & Khilukha, O. (2023). Digitalization of the Ukrainian economy during a state of war is a necessity of the time. *Amazonia Investiga*, *12*(68), 184–194. https://doi.org/10.34069/AI/2023.68.08.17

Vinod, P. P., & Sharma, D. (2021). COVID-19 impact on the sharing economy post-pandemic. *Australasian Accounting, Business and Finance Journal*, *15*(1), 37–50. http://dx.doi.org/10.14453/aabfj.v15i1.4

Viturka, M., Wokoun, R., Krejčová, N., Tonev, P., & Žítek, V. (2013). The regional relationship between quality of business and social environment: harmony or disharmony? *E+M Ekonomie a Management* *16*(2), 22–40. Available at: https://www.ekonomie-management.cz/archiv/search/detail/1024-the-regional-relationship-between-quality-of-business-and-social-environment-harmony-or-disharmony/

Wang, L., Chen, L., & Li, Y. (2022). Digital economy and urban low-carbon sustainable development: The role of innovation factor mobility in China. *Environmental Science* *and Pollution Research*, 29(32), 48539–48557. https://doi.org/10.1007/s11356-022-19182-2

Wang, X., & Wang, Q. (2021). Research on the impact of green finance on the upgrading of China's regional industrial structure from the perspective of sustainable development. *Resources Policy*, *74*, 102436. https://doi.org/10.1016/j.resourpol.2021.102436

Weill, L. (2013). Bank competition in the EU: How has it evolved?. *Journal of international financial markets, institutions and money*, *26*, 100–112. https://doi.org/10.1016/j.intfin.2013.05.005

Wolf, M. J., Emerson, J. W., Esty, D. C., de Sherbinin, A., & Wendling, Z. A. (2022). 2022 Environmental Performance Index (EPI) results. New Haven, CT: Yale Center for Environmental Law & Policy. Available at https://bvearmb.do/handle/123456789/2828

Xie, L., Cheshmehzangi, A., Tan-Mullins, M., Flynn, A., & Heath, T. (2020). Urban entrepreneurialism and sustainable development: A comparative analysis of Chinese eco-developments. *Journal of Urban Technology*, *27*(1), 3–26. https://doi.org/10.1080/10630732.2019.1680940

Zhang, J., Zhao, W., Cheng, B., Li, A., Wang, Y., Yang, N., & Tian, Y. (2022). The impact of digital economy on the economic growth and the development strategies in the post-COVID-19 era: evidence from countries along the "Belt and Road". *Frontiers in public health*, *10*, 856142. https://doi.org/10.3389/Ffpubh.2022.856142

Zhang, S., Li, Y., Liu, Z., Kou, X., & Zheng, W. (2023). Towards a Decoupling between Economic Expansion and Carbon Dioxide Emissions of the Transport Sector in the Yellow River Basin. *Sustainability*, *15*(5), 4152. https://doi.org/10.3390/su15054152

Zhao, T., Zhang, Z., & Liang, S. (2022). Digital economy, entrepreneurship, and high-quality economic development: Empirical evidence from urban China. *Frontiers of Economics in China*, *17*(3), 393. https://doi.org/10.3868/s060-015-022-0015-6

Zheng, J., & Wang, X. (2022). Impacts on human development index due to combinations of renewables and ICTs—new evidence from 26 countries. *Renewable Energy*, *191*, 330–344. https://doi.org/10.1016/j.renene.2022.04.033