UNIVERSITI TEKNOLOGI MARA

JIIICaS 2024
JOHOR INTERNATIONAL INNOVATION, INVENTION COMPETITION AND SYMPOSIUM

PROCEEDINGS OF JOHOR INTERNATIONAL INNOVATION INVENTION COMPETITION AND SYMPOSIUM 2024 (JIIICaS 2024)

"Flourish and Nurturing Sustainable Innovation for a Prosperous Nation"

JIICOS 2024
JOHOR INTERNATIONAL INNOVATION, INVENTION COMPETITION AND SYMPOSIUM

Editorial Board

e ISBN: 978-967-0033-25-9

Published in Malaysia by
Universiti Teknologi MARA Cawangan Johor
Kampus Pasir Gudang
81750 Masai

Preface

In the name of Allah, the Almighty who gives us the enlightenment, the truth, the knowledge and with regards to Prophet Muhammad (peace be upon him) for guiding us to the straight path. We thank to Allah for giving us guidance and strength to write this e-book.

This e-book compiles the extended abstracts that submitted to Johor International Innovation Invention Competition and Symposium 2024 (JIIICaS2024), where JIIICaS2024 is a virtual platform for all creative minds to share and present their invention and innovation. Each abstract gives a brief background on the innovation or project.

We hope that this e-book will help the readers to get to know the innovation done by the students and get some ideas to develop future innovation products.

Foreword Rector

Assalamualaikum warahmatullahi Wabarakatuh, Salam Sejahtera, Salam Malaysia MADANI and Salam UiTM Dihatiku.

In the name of Allah, the Most Gracious, the Most Merciful.

It is a great honor to welcome you to the Johor International Innovation, Invention, Competition, and Symposium 2024 (JIIICaS 2024). This event connects various disciplines, focusing on education and engaging educators, students, researchers, and innovators from all walks of life.

Innovation is not just about ideas; it demands perseverance, creativity, and determination to turn those ideas into reality. The remarkable projects showcased today highlight the dedication and spirit of all participants. Initiatives like this not only explore new technologies but also cultivate skills and leadership among our youth. At Universiti Teknologi MARA (UiTM) Johor Branch, we are fully committed to fostering a dynamic culture of innovation, promoting the commercialization of new products, and encouraging meaningful collaborations with industry and society.

As we celebrate this event, I would like to extend my heartfelt gratitude to all sponsors, judges, the College of Computing, Informatics and Mathematics, UiTM Pasir Gudang Campus as the event organizer, as well as to the researchers and participants for their hard work in making this event a success. Let us continue striving for innovation and excellence. May the ideas presented today inspire us and lay the groundwork for future achievements.

Thank you.

Associate Professor Dr. Saunah Zainon
Rector
Universiti Teknologi MARA (UiTM)
Johor Branch

(A-ST174) FEATURE SELECTION FOR CHRONIC KIDNEY DISEASE CLASSIFICATION MODEL USING CONVOLUTIONAL NEURAL NETWORK

Muhammad Zakwan Zakirin Bin Abd Razak (Zakwan)[1], Dr Izyan Izzati Binti Kamsani (Izyan)[2]

[1]Zakwan
mz.zakirin@graduate.utm.my
[2]Izyan
izyanizzati@utm.my

ABSTRACT

A complicated and common medical disorder, chronic kidney disease (CKD) has a big impact on public health. Early and accurate classification of CKD is crucial for effective management and treatment and late of treatment may lead to serious problem that may cause death. However, determining an accurate classification of CKD is challenging. Using a deep learning as a method for the classification of the targeted disease. In this research, an approach for classifying CKD using a Convolutional Neural Network (CNN) will be proposed. Experiments will be conducted to study the performance of CNN in the classification of CKD. After collecting the dataset of CKD, the raw data will be pre-processed before the action of deep learning can be taken. Several feature selections will be done to target the important attribute that contribute to the CKD. After that, CNN model architecture design will be implemented to the experiment in getting the result for performance evaluation. To evaluate the performance of the result, employment of various performance metrics, including accuracy, precision, recall, and F1 score are recorded. The successful implementation of the proposed CNN model for clinical data classification in CKD can have significant clinical implications, enabling personalized treatments, disease monitoring, and facilitating the development of targeted therapies. This research helped advance the practise of precision medicine for CKD and enhance patient outcomes.

Keywords: Chronic kidney disease, Machine learning, Convolutional Neural Network

INTRODUCTION

The kidneys are vital organs responsible for filtering waste from the blood, maintaining fluid and electrolyte balance, and producing essential hormones. Damage to the kidneys can lead to serious health issues, including chronic kidney disease (CKD), a condition characterized by a gradual loss of kidney function. CKD is a global health concern, affecting over 10% of the population and potentially leading to end-stage renal disease (ESRD), requiring dialysis or transplantation.

CKD is a long-term condition that affects kidney function, commonly linked to aging and prevalent in certain populations. The global prevalence of CKD was 9.1% in 2017, affecting over 700 million people. Machine learning techniques, including CNN, have been explored for CKD prediction, using patient data to develop models that improve early detection and personalized treatment. Various studies have demonstrated the effectiveness of these approaches, with models achieving high accuracy in CKD classification, thus supporting better management of the disease.

This study focuses on improving CKD diagnosis by targeting important features and using Convolutional Neural Networks (CNN), a deep learning technique that leverages clinical data for more accurate disease classification. By enhancing diagnostic accuracy, this approach aims to facilitate earlier intervention, improve patient outcomes, and reduce the burden of CKD.

Classifying chronic kidney disease using CNN has the potential to improve diagnosis accuracy and efficiency, allowing for earlier intervention and improved disease management. The results from this study demonstrate that using Recursive Feature Elimination (RFE) for feature selection yielded the highest accuracy (0.9917) and F1 Score (0.9933), surpassing the performance of models utilizing all features, Chi-Square Test, and Ridge Regression. This suggests that careful feature selection can significantly enhance the performance of machine learning models in CKD classification. Ultimately, this approach aims to improve patient outcomes and reduce the burden of this chronic illness.

METHODS

A research employ needs to get started by gathering materials and data. At this point, the research's references have been found. Usually, information and resources are obtained via the Internet by utilising a browser such as Google Scholar, which provides pertinent content from both the academic and professional domains. Obtaining journals mostly from ScienceDirect and conference papers primarily from IEEE Xplore is the next stage. For this research, the database chosen is chronic kidney disease dataset from UCI Machine Learning Repository. The dataset consists of 400 samples. In this research the performance measurement using convolutional neural network for chronic kidney disease is measured based on accuracy, precision, recall and f1 score. Below is the research framework.

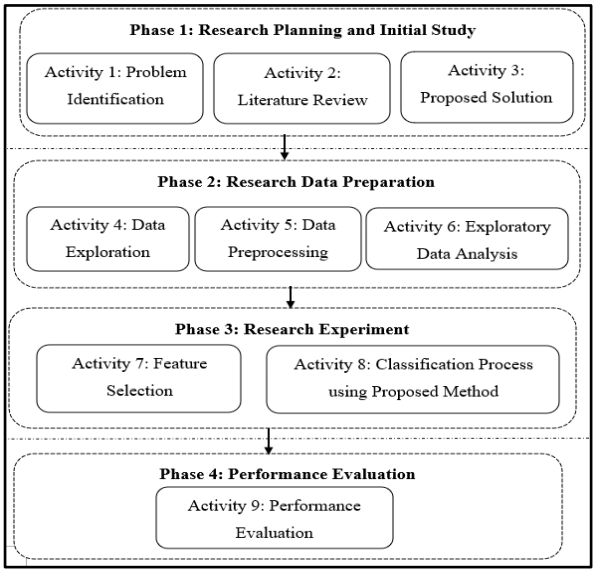

Figure 1 Research Framework

Phase 1: Research Planning and Initial Study

Research planning and conducting an initial study are essential steps in the research process. These stages involve careful preparation and groundwork to ensure the successful execution of the research project. In this phase, there are three actions that need to be initiate to gain information details in specific research. The first action is the identification of the problem for gaining the purpose and objective for doing the

experiment. The next action that needs to be done is doing a literature review based on previous research to get better understanding about machine learning technique that can be use on the classification of CKD model. The third action which is the last action that need to be carry out for the phase is concluding a solution for suitable machine learning technique for the research. Overall, research planning and conducting an initial study are critical steps that set the foundation for a successful research project. It will ensure that the research is well-prepared, feasible, and capable of generating reliable and valid results. Figure below shows the list of activities that can be done on phase 1.

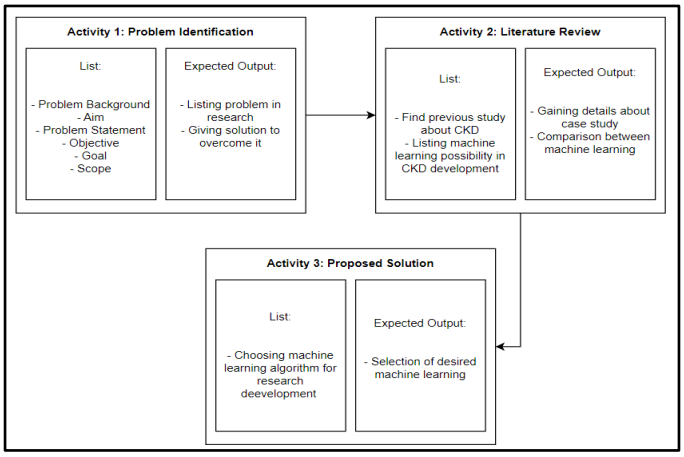

Figure 2 Phase 1

Phase 2: Research Data Preparation

Research data preparation is a crucial step in the research process that involves organizing, cleaning, and transforming raw data into a format suitable for analysis. This stage ensures that the data is accurate, consistent, and ready for meaningful interpretation. The first and foremost step is getting a dataset of chronic kidney disease is an important task for this research experiment so that the guideline will able to follow the following data. The dataset that gets from websites can be considered raw and it need to be clean which can improve the result. Follow to the next action that are need to be taken which is data preprocessing. There are two steps of actions that need to be taken at a same time which are data pre-processing and exploratory data analysis. Data pre-processing encompasses a series of techniques and procedures aimed at improving data quality, addressing inconsistencies, and enhancing the effectiveness of data analysis while exploration of data will help to get an understanding of the data. In the phase 2, it will ensure the data is in a state of accurate, reliable and usable that can improve the accuracy in research development finding. Figure below shows the list of activities that can be done on phase 2.

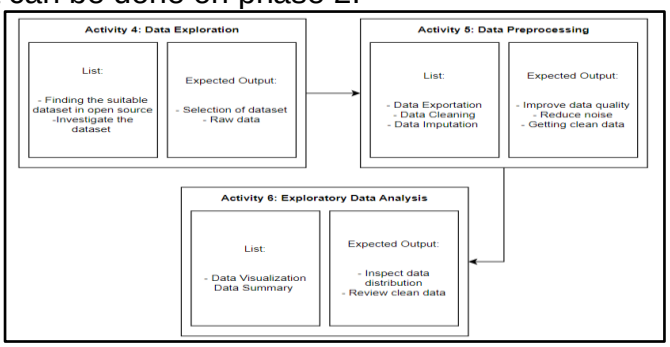

Figure 3 Phase 2

Phase 3: Research Experiment

Research development is an important step for research process that will advancing the research experimentation. This phase helps to get the result for the classification. For the phase 3, feature selection by selecting the most relevant and informative features to improve model performance, reduce overfitting, and enhance interpretability will be conduct. Then, convolutional neural network deep learning algorithm will be implemented across the research. Layers that involved in CNN such as convolutional layer, pooling layer and fully connected layer need to be initiate in etiquette manner. Research development is a dynamic and iterative process that requires implementation of machine learning technique with the proposed data. Outstanding development will help getting a result that will be evaluate in the next phase. Figure below shows the list of activities that can be done on phase 3.

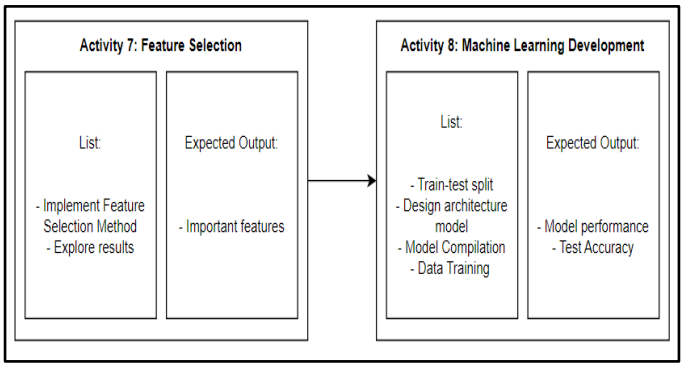

Figure 4 Phase 3

Phase 4: Research Evaluation

Research evaluation is a process that finding the performance of the experiment that are need to be conclude. The last phase, which is phase four will be comparison of the result of the development with other research that using different machine learning technique that has been done. It will help to compare which machine learning technique that is better for classification of chronic kidney disease model. Figure below shows the list of activities that can be done on phase 4.

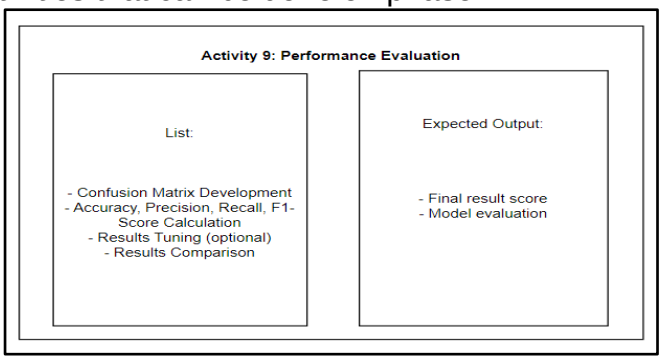

Figure 5 Phase 4

RESULTS AND DISCUSSION

Metrics scores need to be calculated in order to assess how well machine learning models perform. Metrics like recall, accuracy, precision, and F1-score reveal how effectively the model predicts results in comparison to the real data. Precision evaluates the accuracy of positive predictions, or the number of relevant instances selected, whereas accuracy represents the overall percentage of correctly classified instances. Conversely, recall evaluates the model's accuracy in identifying every positive instance. The F1-score is a single metric that measures the overall

performance of the model by balancing precision and recall through a harmonic mean. The dataset is conducted through 4 different method which are All Features, Chi-Square Test, Recursive Feature Elimination and Ridge Regression. The dataset being tested and compared using with Convolutional Neural Network. The results for each experiment was collected based on accuracy, precision, recall and F1 score. Table 1 shows the metrics score of it.

Table 1 Metrics Score

Evaluation Metrics	All Features	Chi-Square Test	RFE	Ridge Regression
Accuracy	0.9833	0.9833	0.9917	0.9833
Precision	1.0	1.0	1.0	0.9740
Recall	0.9733	0.9733	0.9867	1.0
F1 Score	0.9865	0.9865	0.9933	0.9867

1. Result Comparison

The performance of a chronic kidney disease classification model based on Convolutional Neural Networks was assessed using four feature selection techniques: All Features, Chi-Square Test, Recursive Feature Elimination (RFE), and Ridge Regression. Evaluation parameters included accuracy, precision, recall, and F1 score, revealing significant performance variation depending on the feature selection strategy. RFE exhibited the highest accuracy of 0.9917, while the other methods, including using all features, Chi-Square Test, and Ridge Regression, demonstrated an accuracy of 0.9833, indicating similar effectiveness in correct classifications but a slight edge for RFE. Precision was a perfect 1.0 for models using all features, Chi-Square Test, and RFE, but slightly lower for Ridge Regression at 0.9740, suggesting minor false positives. Recall was perfect (1.0) for Ridge Regression but slightly lower for other methods, showing Ridge Regression's ability to identify all chronic kidney disease instances correctly. The F1 score was highest for RFE at 0.9933, balancing precision and recall effectively. Feature selection analysis showed the Chi-Square Test prioritized critical medical conditions and laboratory results, RFE emphasized a broader range of laboratory and demographic features, and Ridge Regression combined essential lab tests and vital signs. 'Specific gravity,' 'albumin,' and 'hypertension' were common across all methods, indicating their importance in predicting chronic kidney disease. The results suggest that while the Chi-Square Test prioritizes specific medical conditions, RFE and Ridge Regression incorporate a broader range of variables, with RFE offering the highest accuracy and F1 score, demonstrating the benefits of a comprehensive feature selection approach in enhancing the CNN model's classification ability. This underscores the significant impact of careful feature selection on the effectiveness of machine learning models for CKD classification, with RFE emerging as the most successful method.

2. Previous Research Result Comparison

In comparing our chronic kidney disease (CKD) classification model using Convolutional Neural Networks (CNN) with previous research, notable differences and similarities in performance metrics emerge. Our CNN model with Recursive Feature Elimination (RFE) achieved an exceptional accuracy of 0.9917, surpassing all other techniques, including those from earlier studies. This highlights RFE's effectiveness in feature selection, enhancing the CNN model's predictive power. While CNN models using all features, the Chi-Square Test, and Ridge Regression demonstrated a consistent accuracy of 0.9833, underscoring the CNN architecture's robustness even without advanced feature selection. This performance is competitive with previous models such as XGBoost and LGBM, which achieved an accuracy of 0.983, and outperforms models like Ada Boost, Stochastic Gradient Descent, and Extra Tree, which achieved accuracies of 0.975. Traditional models like Decision Tree and Naïve Bayes showed lower accuracies of 0.96 and 0.94, respectively, while simpler models like KNN (0.65), ANN (0.60), and SVM (0.93) performed notably worse. The hybrid approach of GA+SVM, with an accuracy of 0.9833, illustrates the benefits of combining feature selection with machine learning models, though it does not match the accuracy of our CNN with RFE. Overall, our findings demonstrate that CNNs, particularly when paired with effective feature selection methods like RFE, offer superior classification performance in CKD prediction, emphasizing the importance of advanced feature selection in enhancing deep learning models. Table 8 shows the result comparison of it.

Table 8 Result Comparison

Method	**Current Research**	**Previous Research**
CNN–All Features	0.9833	-
CNN–Chi-Square Test	0.9833	-
CNN–Recursive Feature Elimination	0.9917	-
CNN–Ridge Regression	0.9833	-
Ada Boost	-	0.975
Decision Tree	-	0.96
XgBoost	-	0.983
CatBoost	-	0.966
KNN	-	0.65
Naïve Bayes	-	0.94
Stochastic	-	0.975
LGBM	-	0.983
Extra Tree	-	0.975
SVM	-	0.93
ANN	-	0.60
GA + SVM	-	0.9833

CONCLUSIONS

The experimentation with feature selection methods for chronic kidney disease (CKD) classification using Convolutional Neural Networks (CNN) provided valuable insights into their effectiveness. Recursive Feature Elimination (RFE) achieved the highest accuracy of 0.9917, surpassing other methods including the use of all features, the Chi-Square Test, and Ridge Regression, which all attained an accuracy of 0.9833. This underscores the crucial role of effective feature selection in enhancing model performance. RFE, Ridge Regression, and the Chi-Square Test each identified distinct feature subsets, with Chi-Square focusing on significant medical conditions and laboratory results, while RFE and Ridge Regression included a broader range of continuous variables and vital signs. Common features like 'specific gravity,' 'albumin,' and 'hypertension' were consistently identified, highlighting their importance in predicting CKD. RFE's superior accuracy and F1 score of 0.9933 reflect its effectiveness in balancing precision and recall, while Ridge Regression, despite perfect recall, showed slightly lower F1 scores and precision compared to RFE. These findings demonstrate that a comprehensive and balanced feature selection approach is essential for optimizing model performance, emphasizing the need for careful selection to enhance predictive accuracy and efficacy in disease classification models.

REFERENCES

Gounden, V., Bhatt, H. and Jialal, I. (2022) Renal Function Tests, National Center For Biotechnology Information. Available at: https://www.ncbi.nlm.nih.gov/books/NBK507821/.

Ching Chun, S. et al. Jui Yi, C, Sheng Yin, C, Chih Chung, S, Javier A., N, Ryo, M, Eisei, N, Emily, S, Yih Ting, C, Cheng Kai, H, Heng Chih, P, Chih Hsiang ,C, Mitchell H., R and Vin Cent, W, (2023) Outcomes associated with acute kidney disease: A systematic review and meta-analysis, eClinicalMedicine. Available at: https://www.sciencedirect.com/science/article/pii/S2589537022004898

Abinet M. Aklilu MD, MPH (2023) Diagnosis of Chronic Kidney Disease and Assessing Glomerular Filtration Rate, Medical Clinics of North America. Available at https://www-sciencedirect-com.ezproxy.utm.my/science/article/abs/pii/S0025712523000342.

Aune, D. et al. (2022) Self-reported chronic kidney disease and the risk of all-cause and cause-specific mortality: Outcome-wide association study of 54 causes of death in the National Health Interview Survey, BMC nephrology. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9055730/.

Goodby, A., Olsen, L., McGinnis, M., (2010) Clinical Data as the Basic Staple of Health Learning: Creating and Protecting a Public Good: Workshop Summary. National Academies Press (US), Available at: https://www.ncbi.nlm.nih.gov/books/NBK54302/.

Agarwal, Vivek. (2015). Research on Data Preprocessing and Categorization Technique for Smartphone Review Analysis. International Journal of Computer Applications. Available at: https://www.researchgate.net/publication/291019609_Research_on_Data_Preprocessing_and_Categorization_Technique_for_Smartphone_Review_Analysis.

Jie Cai, Jiawei Luo, Shulin Wang, Sheng Yang (2018). Feature selection in machine learning: A new perspective, Neurocomputing. Available at: https://www.sciencedirect.com/science/article/abs/pii/S0925231218302911