

Hybrid Diagnostic Model for Kidney Disease Prediction Using Data Mining Techniques

Guli Usa John, Yusufu Musa Malgwi, Ibrahim Adamu and Ezekiel Peter

Department of Computer Science, Faculty of Physical Science Modibbo Adama University, Yola, Nigeria.

Corresponding author: usaguli3@gmail.com

Abstract

This paper proposes and evaluates a hybrid diagnostic model for kidney disease prediction using data mining techniques, as a potential solution to the current problems facing existing models. The model is based on a combination of Artificial Neural networks, and Support Vector Machines algorithms, trained on the symptoms and patient-reported data. The dataset contains 400 instances which are based on 25 attributes, retrieved from University of California, Irvine machine learning repository chronic_Kidney_Disease Dataset. WEKA toolkit was used to preprocess dataset, apply the data mining algorithms, analysis and evaluation. The evaluation of the model using a 10-fold cross-validation technique shows that the hybrid model outperforms both individual ANN and SVM algorithms in terms of accuracy and stability. Also, it shows that the hybrid of ANN-SVM has the lowest MAE, RMSE, and RRSE in classifying the CKD dataset, followed by ANN having lower MAE compared to ANN, but ANN had lower RMSE, RAE, and RRSE than the SVM. The model described in this paper is an efficient tool for predicting kidney disease, as it can be used to diagnose a variety of kidney diseases and can potentially reduce the number of unnecessary tests that are currently conducted to diagnose kidney disease. In addition, the model can be used to ensure that patients receive the most appropriate medical screenings and treatments for kidney diseases. In conclusion, the proposed hybrid diagnostic model for kidney disease prediction using data mining techniques has shown that it is a viable and promising alternative to existing methods.

Keywords: Artificial Neural Network, Chronic Kidney Disease, Prediction, Data Mining Techniques, Machine Learning, Support Vector Machine.

Introduction

The kidney is an important organ in the human body that filters blood or removes metabolic wastes and surplus fluid from the circulatory system through blood plasma and excretes them in the urine. It also controls the electrolyte and fluid balances throughout the body. Kidney disorders are referred to as silent killers because up to 90% of renal function decline can occur without any symptoms or problems (Bala and Kumar, 2014). According to Bikbov, Perico, and Remuzzi (2018), renal failure was the greatest cause of mortality among people residing in contemporary civilization in 2016, killing 753 million people worldwide due to chronic kidney disorders (CKD), including 336 million men and 417 million women. The demand for quick, inexpensive, and reliable illness diagnosis methods is currently the

biggest challenge facing the health industry (Bali, Garba, & Ahmadu, 2021). Early kidney impairment diagnosis may aid in correction, even though this is not always practicable. Data mining is a technology for extracting implicit, unknown, and possibly beneficial information from data. It is often referred to as knowledge discovery in databases (Lashari, Ibrahim, Senan & Taujuddin, 2018). The healthcare sector is one of the most significant industries adopting data mining to create a lot of data about patients, illnesses, hospitals, medical equipment, and treatment costs, among others. The goal of data mining is to find and extract hidden information, such as different patterns, regularities, and anomalies, from massive datasets. Data mining is a collection of automated procedures (Abdelaziz, Salama, Riad & Mahmoud, 2019). To support the work of medical staffs, data mining is crucial for the analysis of healthcare data in order to uncover hidden patterns and patients' early illnesses (Smecca, Numata, Deretzis, Pellegrino, Boninelli *et al.*, 2016).

Statement of the Problem

Medical practitioners have labelled several diseases including chronic kidney disease as deadly diseases that requires urgent attention (Olatunde, Omotosho and Akanbi, 2019). Experts encounter a lot of difficulties in reaching conclusion on several deadly diseases due to symptoms similarities (Cherian and Bindu, 2017). The advancement in technology and promising contribution of machine learning algorithm shows that the medical section can get more reliable support to save more lives. The majority of healthcare-related diseases, according to technological research statistics, use specialized data mining techniques that don't produce the best outcomes due to factors like decision-making that isn't very accurate, a lack of expertise, challenges with knowledge upkeep, and time-dependent performance. Research by Vijayarani, Dhayanand, and Phil (2015) just but few have identified the need for hybrid techniques for better performance and minimized execution time. Due to these issues, it is necessary to apply hybrid data mining to provide the diagnostic help mechanism.

The Objectives of the Study

The objectives include:

- to extract and preprocess the dataset of symptoms for feature extraction and to reduce the irrelevant features from set features,
- to develop the Hybrid Model using SVM and ANN to detect chronic kidney disease, and
- to simulate the Hybrid Model using the chronic kidney disease dataset.

The primary goal of this study is to develop and apply a hybrid diagnosis model for categorizing kidney disease. This advocated that ANN algorithm need to be enhanced to minimize the execution time which is in line with the aim of this study. To do this, data from the selected sources was preprocessed, three machine learning classification approaches

were used to predict the disease. The Hybrid Model was developed, simulated and validated.

The importance of the hybrid diagnostic model approach is it has the potential to revolutionize kidney disease management by offering improved accuracy, personalized predictions, and cost-effective healthcare solutions, ultimately leading to better patient outcomes and enhanced public health strategies.

Literature Review

Medical practitioners have labelled several diseases including chronic kidney disease as deadly diseases that require urgent attention (Olatunde, Omotosho & Akanbi, 2019). Lashari, Ibrahim, Senan and Taujuddin (2018) explained that diabetes is also a major cause of chronic kidney disease. According to Bala and Kumar (2014) kidney failure such as Kidney stones, Glomerulonephritis, polycystic kidney and urinary tract infections can occur when your body becomes overloaded with toxins.

Experts encounter a lot of difficulties in reaching conclusion on several deadly diseases due to symptoms similarities (Cherian & Bindu, 2017). Data mining techniques are employed to support the effort of medical personnel in reaching early conclusion and making accurate decision in order to save the lives of the patients (Cherian & Bindu, 2017). According to Yasin and Prasad Rao (2018) data mining is a method that allows us to examine huge sets of data and understand their meaning after extraction. From clinical and diagnostic data, the data mining and healthcare industries have developed a number of trustworthy early detection systems and other associated healthcare systems (Jothi and Husain, 2015). For clinical decision-making, data mining can extract fresh biomedical and healthcare information (Parvathi and Rautaray, 2014).

Numerous mathematical methods, including decision trees, neural networks, logistic regression, support vector machines, evolutionary algorithms, and Bayesian networks, are used in the classification process. The dataset can be used by classification software to learn how to forecast future events. The most popular method of identification, prediction, and optimization in the healthcare sector is classification (Reddy, Mandakini and Radhika, 2020). The data mining model analyzes the data using data mining methods after obtaining it from the mining structure. Information defining the data source is stored in the mining structure. In the process of developing a model, methods like decision trees, neural networks, and logistic regression can be used (Reddy, Mandakini and Radhika, 2020). Artificial neural networks (ANN) are biologically inspired computer programs designed to simulate the manner in which the human mind works. The ANN have the capacity to interact with their environment, learn from it, and self-organize their structure. However, it is incapable of logical thought or justification (Garba & Bali, 2021).

Several study have explored the decision tree method to analyze clinical data. In their separate work Zolbanin, Delen, and Zadeh (2015) employed the decision tree algorithm. With regards to accuracy, precision, and execution time for the prediction of chronic renal

disease, Devika, Avilala, and Subramaniaswamy (2019) compare the performance of Naive Bayes, K-Nearest Neighbour (KNN), and Random Forest classifiers. The outcome revealed that the performance of Random Forest classifier is finest than Naive Bayes and KNN. Makvana and Kotak (2019) presented various heart disease prediction techniques and highlighted the need to compare classification algorithm such as SVM and ANN for the prediction of cardiovascular disease. Rady and Anwar (2019) presented the use of efficient data mining techniques is shown to reveal and extract hidden information from clinical and laboratory patient data, which can be helpful to assist physicians in maximizing accuracy for identification of disease severity stage. Nagendra, Ussenaiah and Rajasekhar (2020) considered Neural Network (NN), Support Vector Machines (SVM), eXtreme Gradient Boosting (XGB), Random Forest (RF), Linear Discriminant Analysis (LDA) and Enhanced Gradient Boosting (EGB) is Data Mining Classification Model which is extended from XGB in the prediction of heart disease.

Devika *et al.* (2019) give a comparative analysis of a different classifier to predict kidney illness. The results of Naive Bayes, KNN, and Random Forest showed that the accuracy of NB was 99.635%, KNN was 87.78%, and Random Forest was 99.844%. Vijayarani, Dhayanand, and Phil (2015) compared the performance of the two algorithms on the basis of accuracy and execution time using several medical labs, centres and hospitals, dataset created for analysis of kidney disease. They observed that the performance of the artificial neural network was better than that of support vector machine algorithm, however, ANN takes longer training time and not able to generalize from limited data. Conversely, SVM classifier classifies the data with minimum execution time and can map data to an infinite dimensional space. The single linear technique in the diagnosis of kidney disease has been found to be less effective to derive expected result.

Parvathi and Rautaray (2014); Reddy, Mandakini and Radhika (2020) also observed that data mining algorithms can be enhanced by hybridizing algorithms, as a single algorithm may not be accurate for classified sets of data. Researches by Vijayarani, Dhayanand, and Phil (2015) have identified the need for hybrid techniques for better performance and minimized execution time.

Materials and Methods

The aim of this study is to compare the performance of the two algorithms on the basis of accuracy and execution time. The dataset used was collected from several medical labs, centres and hospitals, created for analysis of kidney disease. The dataset contains five hundred and eighty-four (584) instances and six (6) attributes.

Description of the Dataset

The data used in this study was collected by downloading it from University of California, Irvine (UCI) machine learning repository chronic_Kidney_Disease (CKD) Dataset. The data included the chronic kidney disease data of 400 patients and has 25 attributes. The dataset

contains 400 instances which are based on 25 attributes. The dataset contains 24 features, divided into 11 numerical traits and 13 categorical features, in addition to the class features, such as "CKD" and "not-CKD," for classification. The dataset characteristics are multivariant. It was donated in 3rd of July 2015 by Dr. P. Soundarapandian from Apollo Hospitals, Managiri, India (UCI, 2021).

Description of the Proposed Hybrid Model

The architecture of how single and hybrid technology was developed with the combination of properties of artificial neural network and support vector machine data mining techniques is as presented. The hybrid data mining techniques was utilized for predication of kidney disease diagnosis based on given input symptoms in medical diagnosis. The methodology is applied to data collected from University of California Ir-vine (UCI) CKD dataset. The proposed model uses ANN and SVM data mining algorithm each on kidney disease data and check for technique accuracy, then apply the hybrid data mining technique on kidney disease diagnosis data to test its accuracy. This analysis is to find which data mining technique is best for predication and compared with the hybrid data mining technique. The hybridized data mining techniques is very helpful to physicians to make a faster and more accurate decisions.

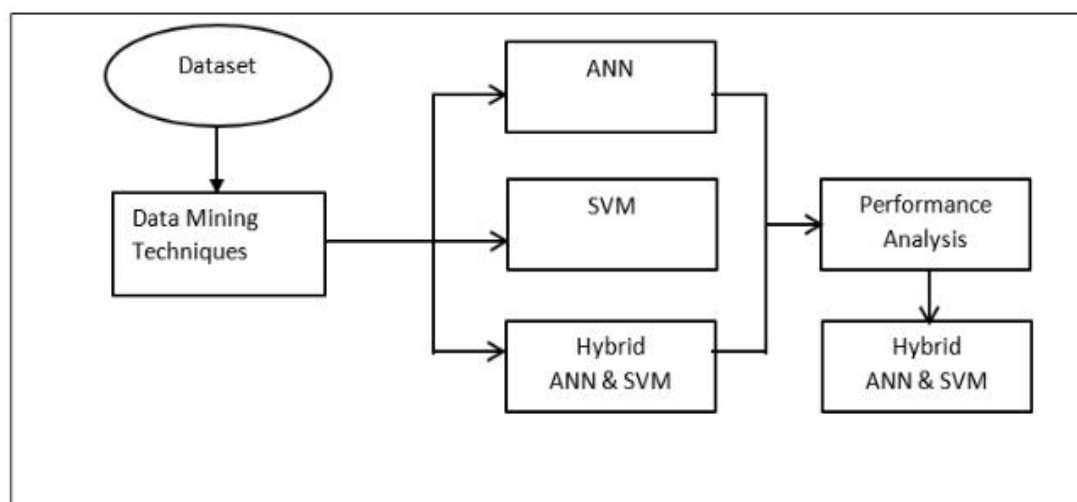


Figure 1: Shows the proposed approach of using hybrid technique and compares the single technique and hybrid technique.

Dataset Pre-processing

The dataset included outliers and noise. Therefore, different data pre-processing approaches have been followed to prepare the dataset for the model. To build the model, the dataset was cleaned from the missing values for transformation of data. The estimation of the missing values and noise elimination, such as outliers, normalization, and imbalanced data checks, were included in the preprocessing step because certain measures may be lost

when patients are being tested, leading to missing values. The missing values of nominal features have been filled by mode. The mean has been used to fill in the missing values for numerical features. Figure 2 shows that there is no missing data available in this dataset after preprocessing.

Model Development

The process of creating the model follows the completion of data preparation and handling the unbalanced dataset. The ANN, SVM, and hybrid algorithms have been employed after preprocessing and feature selection. In this study, Waikato Environment for Knowledge Analysis (WEKA) version 3.8.2 was used. WEKA is a powerful open source data mining platform, with a toolkit to machine learning and DM was used to apply the data mining algorithms. Weka includes several tools for regression, classification, clustering, association rules, visualization, and data pre-processing software. The Experiments was carried out on a 64-bit Windows 10 Pro operating system, with 4 GB of RAM and an Intel (R) Core (TM) i5 CPU at 2.30GHz per core. Due to the iterative nature of the experiments and resultant processing power required, the java heap size for WEKA-3-9.5 was set to 1024 MB.

Mathematical Description of Hybrid ANN and SVM Model

The hybrid ANN-SVM model is a combination of Artificial Neural Networks (ANN) and Support Vector Machines (SVM). The aim of the hybrid model is to improve the accuracy and generalization of the SVM model by using the powerful learning capabilities of ANN. Here is the mathematical derivation for the hybrid ANN-SVM model. The mathematical representation of SVM-NN can be expressed as follows:

First, we define the SVM model as:

$$y = \text{sign} \left(\sum_{i=1}^n \alpha_i y_i K(x_i, x) + b \right) \quad (1)$$

where α_i represents the Lagrange multipliers to satisfy the equalities,

y_i represents the class labels,

$K(x_i, x)$ = is the kernel function, and x_i, x are n dimensional inputs,
 b is the bias term.

Next, the ANN model is defined as:

$$y = f \left(\sum_{i=1}^n w_i x_i + b \right) \quad (2)$$

where $\mathbf{w}_i = \{w_1, w_2, \dots, w_n\}$ represents the weights vectors,

$\mathbf{x}_i = \{x_{1i}, x_{2i}, x_{3i}, \dots, x_{ni}\} \in R^n$ represents the input features,

b is the bias term, and

f is the sigmoid-shaped nonlinear activation function.

To combine the two models, the output of the SVM is used as input to the ANN given by:

$$y = f\left(\sum_{i=1}^n w_i \left(\sum_{j=1}^m \alpha_j y_j K(x_j, x_i) + b_i\right) + b\right)$$

(3)

where m is the number of support vectors in the SVM model.

b_i are biases vectors, and

y is the final output.

To reduce the error between the output of the ANN and the true labels, the weights and biases b_i of the ANN are changed during training. Overall, the hybrid ANN-SVM model combines the strengths of both models to improve the accuracy and generalization of the SVM model.

Simulation Process

The simulation tool that was used in this project was WEKA. WEKA is a comprehensive piece of software that enables you to pre-process large amounts of data, use several machine learning techniques on large amounts of data, and compare different results. The experimental flow is based on the following steps.

Step 1: Selected kidney disease dataset is retrieved and stored in local repository

Step 2: the WEKA Simulation tool is used to load and pre-process the dataset.

Step 3: Support Vector Machine is triggered and applied on chronic kidney disease dataset

Step 4: Artificial Neural Network is triggered and applied on chronic kidney disease dataset

Step 5: Hybrid of SVM and ANN is formulated using stacking method and applied on chronic kidney disease dataset.

Step 6: The result obtained in Step 3, Step 4 and Step 5 are compared and analyzed.

Performance Evaluation Metrics

The evaluation metrics that were used to evaluate the performance of the classifiers are; Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Relative Absolute Error (RAE), and Root Relative Square Error (RRSE) as shown by the equations 4 to 7.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

(4)

Where;

n = sample size

y_i = actual value

$\hat{y}_i$ = predicted value

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

(5)

$$RAE = \frac{\frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|}{\frac{1}{n} \sum_{i=1}^n |y_i - \bar{y}|}$$

(6)

Where

$\bar{y}$ is the mean of the n actual value

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

(7)

Following the computation of the missing values, it is necessary to find the significant features that strongly and favorably correlate with those traits critical for disease diagnosis. Extracting features that are relevant and eliminates useless features makes prediction algorithm a robust diagnostic model. In this study, we used ANN, SVM, and the combination of ANN-SVM.

Results and Discussion

Table 1: Statistical analysis of ckd dataset (numerical features)

Features	Mean	Standard deviation	Max	Min
Age	51.483	17.21	90	2
Blood glucose random	148.037	76.583	490	22
Serum creatinine	3.072	4.512	76	0.4
Blood pressure	76.469	13.756	180	50
Blood urea	57.426	49.987	391	1.5
Potassium	4.627	2.92	47	2.5
Packed cell volume	38.884	8.762	54	9
Sodium	137.529	9.908	163	4.5
Hemoglobin	12.526	2.815	17.8	3.1
White blood cell count	8406.12	2823.35	26400	2200
Count of red blood cells	4.707	0.89	8	2.1

University of California, Irvine machine learning repository chronic_Kidney_Disease Dataset (2021)

Table 1 displays weights of all features that are selected. The 12 most significant features—wc, bgr, bu, sc, pcv, al, haem, age, su, htn, dm, and bp—have the highest scores and were used to predict CKD. One can see that whereas bp has the lowest score—80.02—wc has the highest—12733.72. Bgr recorded the second-highest score with a 2428.327. Sc and pcv. RBC records the highest weight at 0.4551, while BP records the lowest weight at 0.01584.

Result of Artificial Neural Network Algorithm

The data is partitioned into the two classes of “ckd” and “notckd”. The objective is to separate ckd and notckd. First a classifier was constructed using the training data and then testing data was tested with the constructed classifier to classify the data into ckd” and “notckd. Figure 2 shows the simulation when ANN algorithm is applied on the pre-processed dataset.

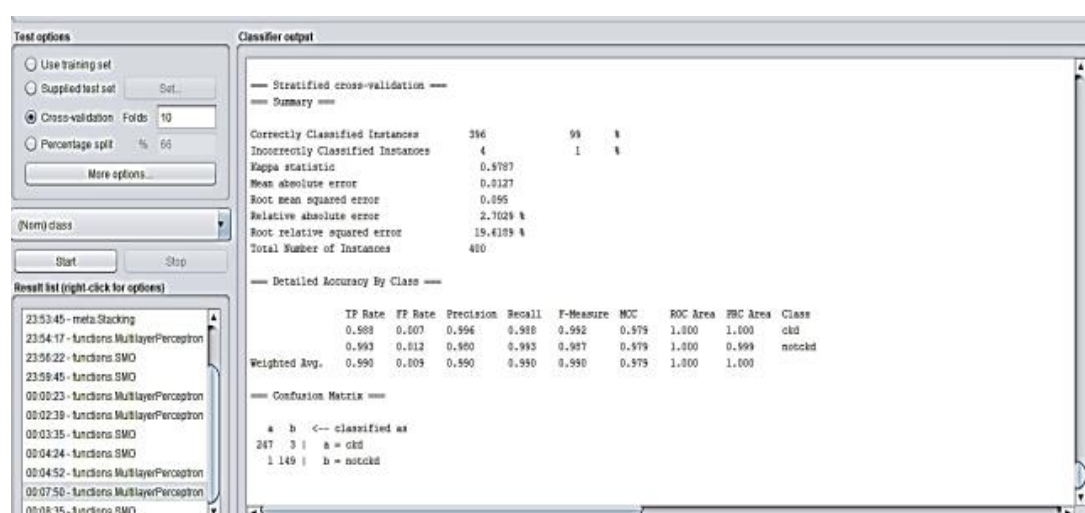


Figure 2: Presents the simulation when ANN algorithm is applied on the pre-processed CKD dataset

Result of Support Vector Machine Algorithm

After employing the Support Vector Machines (SVMs) algorithm the following results were obtained as shown by the SVM as in Figure 2.

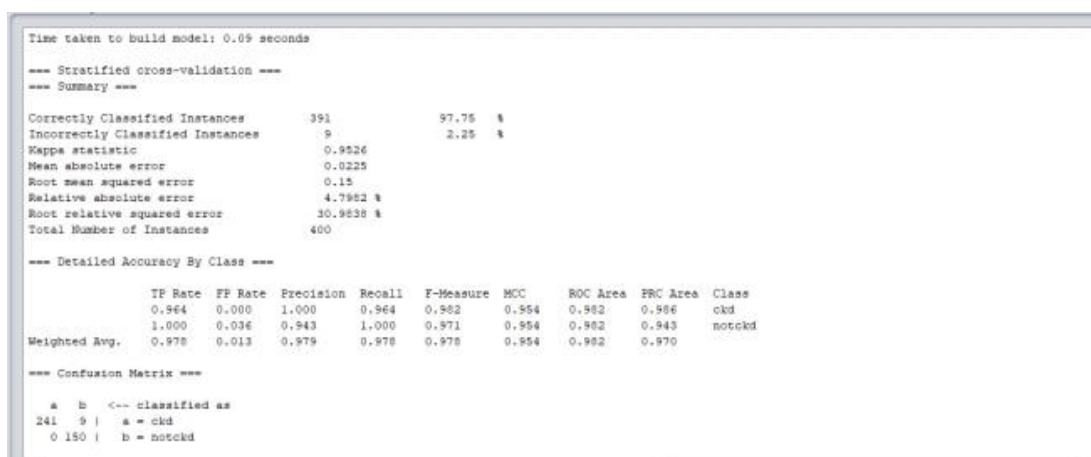


Figure 3: presents the simulation when SVM algorithm is applied on the pre-processed CKD dataset

Hybrid of ANN-SVM Algorithm Experiment Result

This section describes the results of the newly constructed predictive model by means of statistical score metrics and diagnostic visualizations on CKD dataset. Using the hybridized Artificial Neural Network (ANN) and Support Vector Machine (SVM) to form ANN-SVM for detecting chronic kidney disease, the results obtained is as shown in figure 4.

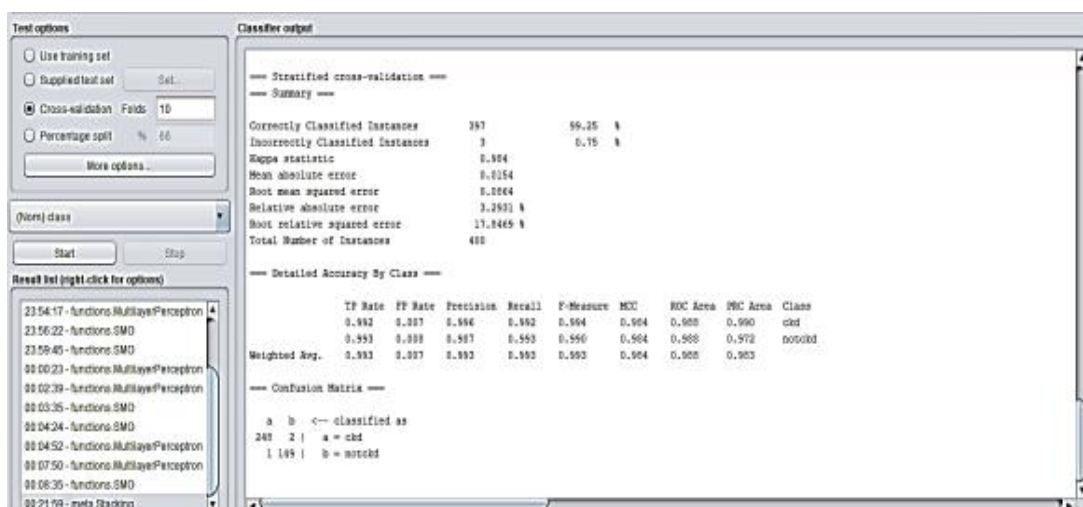


Figure 4: presents the simulation when hybrid ANN-SVM algorithm is applied on the pre-processed CKD dataset

The sample screenshots of the experimental result to determine the percentage correct, percentage incorrect, Kappa statistic, MAE, RMSE, RAE and RRSE are as follows:

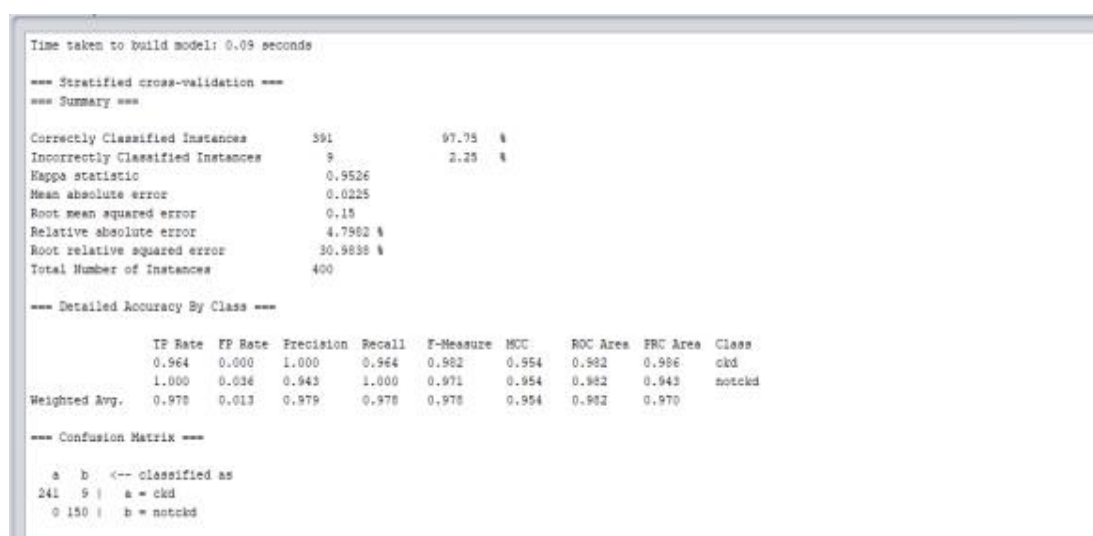


Figure 5: presents the SVM simulation result to test Kappa Statistic on CKD dataset

Comparison of ANN, SVM and Hybrid of ANN-SVM

The comparison between ANN, SVM and ANN-SVM in terms of Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Relative Absolute Error (RAE), and Root Relative Square Error (RRSE) as shown in table 2. It shows the Summary of ANN, SVM and ANN-SVM and compares the three algorithms.

Table 2: Showing the result for SVM, ANN and hybrid ANN-SVM algorithms

Parameters	Algorithm		
	SVM Values	ANN values	Hybrid ANN-SVM values
MAE	0.0170	0.0126	0.0070
RMSE	0.0485	0.0694	0.0415
RAE	3.4028	2.8933	1.1985
RRSE	10.0218	14.3322	8.5708

Total No. of 400 instances in each case.

Table 2 provides the performance of classification algorithms on CKD datasets in terms of MAE, RMSE, RAE, and RRSE. In comparison, it shows that the hybrid of ANN-SVM has the lowest MAE, RMSE, RRSE in classifying the CKD dataset, followed by ANN. Lastly, SVM has lower MAE compared to ANN, but ANN had lower RMSE, RAE, RRSE than the SVM. From all of the results, it can be seen that the hybrid of ANN-SVM model was the best in the detection of kidney disease compared to ANN and SVM in terms of the selected evaluation metrics MAE, RMSE, RAE, and RRSE. It can be seen that with hybrid technique the improvement has occurred in every measure in the case of both the ANN and SVM.

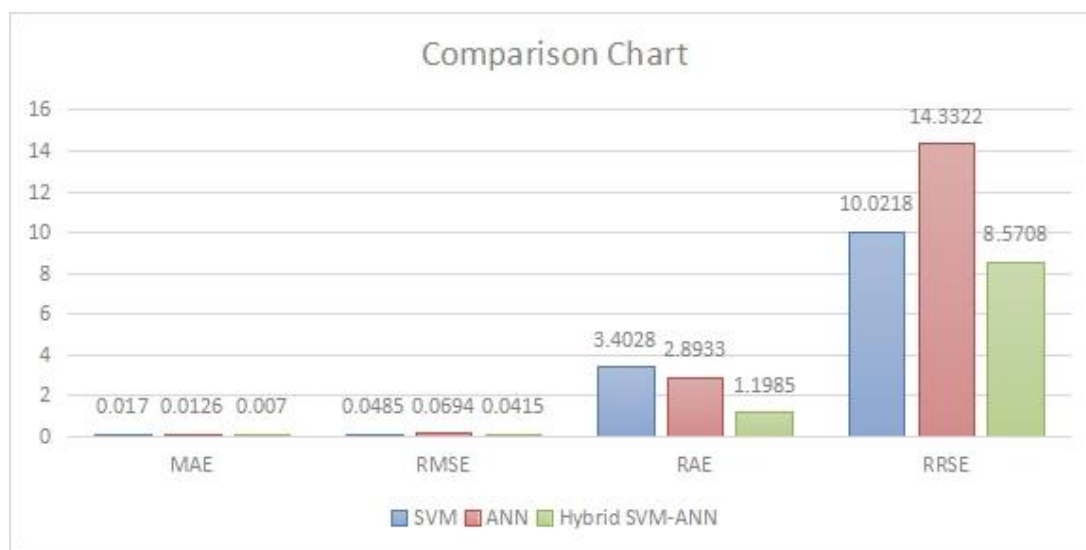


Figure 6: Shows the Comparisons between measured MAE, RMSE, RAE, and RRSE of the three models

Figure 6 shows the RRMSE of the comparative data mining models SVM), ANN, and the proposed hybrid model using the CKD dataset. The proposed model outperforms the individual approaches with a minimum RRSE, MAE, RMSE and RAE giving evidence that the proposed hybrid model showed significant improvement.

Model Performance Evaluation

Several frequently utilized model score metrics were employed to evaluate the performance of the three models. The Evaluation was based on Kappa statistics in addition to root mean square error (RMSE), mean absolute error (MAE), RAE and RRSE with performance parameters during cross-validation (10 folds).

Table 3: Shows the results of model evaluation by using ANN, SVM and hybrid method for CKD on the selected attributes

Parameters	Algorithm		
	SVM Values	ANN values	Hybrid ANN-SVM values
MAE	0.0170	0.0126	0.0070
RMSE	0.0485	0.0694	0.0415
RAE	3.4028	2.8933	1.1985
RRSE	10.0218	14.3322	8.5708
Kappa statistics	0.9526	0.9787	0.9840

From table 3, compared to SVM, ANN had Square Error (RMSE), Relative Absolute Error (RAE) and Root Relative Squared Error (RRSE) and MAE lower than the SVM. In addition, the Kappa statistic value of ANN is 0.9787, which means that the prediction model was

statistically significant. The hybrid techniques have achieved better results than the SVM and the ANN. This support the Yasin and Rao (2018) that utilization of more than one data mining techniques can show accurate results for comparing the performance in the prognosis of heart related disease. However, in terms of the kappa statistics, hybrid techniques have the highest kappa statistics value of 0.984 than the individual techniques. A kappa-statistic is the measure of interpreter agreement when there are two unique raters and two or more ratings and that the value of 1 means that there is perfect agreement between the raters. Therefore, the hybrid method is the improved performance compared to the individual techniques. However, since in most cases, any value of kappa statistic above 0.7 is considered to be very good agreement, it shows that all the techniques achieved a very good result in terms of kappa statistics.

Conclusion

The majority of healthcare-related diseases, according to technological research statistics, used specialized data mining techniques with time-dependent performance and certain amount of errors. Support Vector Machine (SVM) and Artificial Neural Network (ANN) are two of the most known machine learning techniques. Both methods are beneficial and have a history of producing outstanding results across a range of industries. SVM, which has demonstrated the highest accuracy in prior studies, has been contrasted to ANN, which has been claimed to more accurately predict CKD. Compared with other Classifier that also applied chronic kidney disease as a dataset, both ANN-SVM and ANN classifier showed superior performance in ckd and not ckd prediction. However, in terms of Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Relative Absolute Error (RAE), and Root Relative Square Error (RRSE) the hybrid of ANN-SVM showed promising results. This shows that the hybrid approach is more effective than only one approach. Also, an approach that utilizes more than one algorithm could be enhanced and implemented. The novelty of this study are; this study provided a proposed hybrid approach for the detection of kidney disease with reduced error compared to the individual approaches, based on data mining. In the future the researchers would consider applying deep learning methods to detect CKD in diagnosis of the CKD patients.

Recommendations

Based on the results obtained we recommend that;

- healthcare professionals and medical researchers, implement the hybrid model as part of their healthcare information systems to enhance their diagnostic capabilities and support evidence-based decision-making
- Kidney specialists to use the model to aid in the early detection and accurate prediction of kidney diseases in patients, leading improved patient outcomes.

- Data scientists working in the field of nephrology and healthcare analytics to collaborate with healthcare professionals to develop, optimize, and validate the model.
- Patients to benefit from the hybrid diagnostic model by receiving earlier and more accurate diagnoses.

References

- Abdelaziz, A., Salama, A. S., Riad, A. M., & Mahmoud, A. N. (2019). A machine learning model for predicting of chronic kidney disease based internet of things and cloud computing in smart cities. *Security in smart cities: models, applications, and challenges*, 93-114.
- Bala, S., & Kumar, K. (2014). A literature review on kidney disease prediction using data mining classification technique.
- Bali, B., & Garba, E. J. (2021). Neuro-fuzzy approach for prediction of neurological disorders: a systematic review. *SN Computer Science*, 2(4), 307.
- Bali, B., Garba, E. J., & Ahmadu, A. S. (2021). Adaptive Neuro Fuzzy Inference System for Diagnosis of Stimulant Use Disorders.
- Bikbov, B., Perico, N., Remuzzi, G., & GBD Genitourinary Diseases Expert Group. (2018). Disparities in chronic kidney disease prevalence among males and females in 195 countries: analysis of the global burden of disease 2016 study. *Nephron*, 139, 313-318.
- Cherian, V., & Bindu, M. S. (2017). Heart disease prediction using Naive Bayes algorithm and Laplace Smoothing technique. *Int. J. Comput. Sci. Trends Technol*, 5(2), 68-73.
- Devika, R., Avilala, S. V., & Subramaniaswamy, V. (2019, March). Comparative study of classifier for chronic kidney disease prediction using naive bayes, KNN and random forest. In *2019 3rd International conference on computing methodologies and communication (ICCMC)* (pp. 679-684). IEEE.
- Jothi, S., & Anita, S. (2012). Data mining classification techniques applied for cancer disease—a case study using Xlminer. *International Journal of Engineering Research & Technology*, 1(8).
- Lashari, S. A., Ibrahim, R., Senan, N., & Taujuddin, N. S. A. M. (2018). Application of data mining techniques for medical data classification: a review. In *MATEC Web of conferences* (Vol. 150, p. 06003). EDP Sciences.
- Makvana, A., & Kotak, D. (2019). Comparative Analysis of Data Mining Classification Techniques for Cardiovascular Disease Prediction. *International Journal of Engineering Research & Technology (IJERT)*, 8(11), 789-792.
- Nagendra, K. V., Ussenaiah, M., & Rajasekhar, N. (2020, March). Design and Development of EGB Classification Model for predicting Heart Diseases. In *2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)* (pp. 359-366). IEEE.
- Olatunde, Y., Omotosho, L., & Akanbi, C. (2019). Comparison of adaboost and bagging ensemble method for prediction of heart disease. *Heart*, 50, 17.
- Parvathi, I., & Rautaray, S. (2014). Survey on data mining techniques for the diagnosis of diseases in medical domain. *International Journal of Computer Science and Information Technologies*, 5(1), 838-846.
- Rady, E. H. A., & Anwar, A. S. (2019). Prediction of kidney disease stages using data mining algorithms. *Informatics in Medicine Unlocked*, 15, 100178.
- Reddy, R. P., Mandakini, C., & Radhika, C. (2020). A Review on Data Mining Techniques and Challenges in Medical Field. *International Journal of Engineering Research & Technology (IJERT)*, 9(8), 329-333.

- Smecca, E., Numata, Y., Deretzis, I., Pellegrino, G., Boninelli, S., Miyasaka, T., ... & Alberti, A. (2016). Stability of solution-processed MAPbI₃ and FAPbI₃ layers. *Physical Chemistry Chemical Physics*, 18(19), 13413-13422.
- Vijayarani, S., Dhayanand, S., & Phil, M. (2015). Kidney disease prediction using SVM and ANN algorithms. *International Journal of Computing and Business Research (IJCBR)*, 6(2), 1-12.
- Yasin, S. A., & Prasad Rao, P. V. R. D. (2018). Analysis of single and hybrid data mining techniques for prediction of heart disease using real time dataset. *International Journal of Engineering and Technology (UAE)*, 7(2), 97-99.
- Zolbanin, H. M., Delen, D., & Zadeh, A. H. (2015). Predicting overall survivability in comorbidity of cancers: A data mining approach. *Decision Support Systems*, 74, 150-161.