Indonesian Journal of Electrical Engineering and Computer Science
Vol. 41, No. 3, March 2026, pp. 966~976
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v41.i3.pp966-976

Enhanced prediction of chronic kidney disease onset through machine learning techniques

Samuel John Parreño, Maria Cristine Joy Anter
Mathematics Division, University of Mindanao, Digos College, Digos City, Philippines

Article Info

Article history:

Received May 10, 2024
Revised Nov 13, 2025
Accepted Dec 13, 2025

Keywords:

Chronic kidney disease
k-nearest neighbors
Machine learning
Stacking model
Support vector machines
XGBoost

ABSTRACT

Chronic kidney disease (CKD) is a global health concern that often progresses silently to severe complications. This study aims to enhance CKD prediction using machine learning models: support vector machines (SVM), extreme gradient boosting (XGBoost), k-nearest neighbors (k-NN), and a stacking model. The dataset, sourced from the UCI machine learning repository, includes clinical and demographic attributes from 200 patients. After preprocessing, the final dataset comprised 161 samples and 143 features. SVM achieved perfect classification performance with 100% accuracy, precision, and recall. XGBoost followed closely with an accuracy of 97.44% and a kappa statistic of 0.9451. The k-NN model delivered strong performance, achieving 92.31% accuracy. The stacking model outperformed all individual models, achieving perfect accuracy. The models demonstrated high sensitivity and specificity, indicating their effectiveness in distinguishing CKD from non-CKD cases. These findings emphasize the potential of machine learning in CKD diagnosis. Early detection can lead to improved clinical outcomes by enabling timely interventions and personalized treatment strategies. Future research should emphasize comprehensive feature engineering and larger, more diverse datasets to improve predictive accuracy and generalizability. Incorporating machine learning models in nephrology could significantly advance CKD detection and management.

Corresponding Author:

Samuel John Parreño
Mathematics Division, University of Mindanao, Digos College
Digos City, Philippines
Email: samueljohn_parreno@umindanao.edu.ph

1. INTRODUCTION

Chronic kidney disease (CKD) is increasingly recognized as a global public health concern, affecting millions of people worldwide. It serves as a critical determinant of poor health outcomes, including premature mortality. CKD often progresses silently and insidiously, making the identification and treatment of its early stages crucial for improving patient outcomes. Its complications can be severe, including hypertension, anemia, bone disease, and a significantly increased risk of cardiovascular diseases and end-stage renal disease, necessitating dialysis or kidney transplantation [1]–[3]. The prevalence of CKD is notably higher in regions with limited access to primary health care and where educational and economic disadvantages prevail. This is particularly evident in rural areas of developing countries where medical resources are scarce, and awareness about the disease is lacking [4]–[6]. The complexity of CKD, combined with its asymptomatic nature in early stages, poses significant challenges for healthcare systems, often resulting in delayed diagnoses and suboptimal management.

Modern healthcare has begun to leverage advancements in data analytics and machine learning to address the challenges associated with early detection and ongoing management of chronic diseases like CKD. Machine learning offers powerful tools for sifting through large volumes of data to detect patterns that may indicate early stages of kidney disease [7], [8]. These patterns, often imperceptible to human analysts, can include subtle changes in kidney function over time or correlations between various risk factors and the progression of CKD [9], [10]. Despite significant advancements in machine learning applications for CKD prediction, there remains a gap in the implementation of these technologies in real-world clinical settings, particularly in under-resourced areas. Additionally, previous studies have often focused on single models, whereas our approach integrates multiple models and a stacking technique to potentially enhance predictive accuracy. This gap in technology application highlights an urgent need for more targeted research and development efforts that can bring the benefits of machine learning to the forefront of kidney disease prevention and treatment. To this end, our research focuses on the application of sophisticated machine learning algorithms to analyze the risk factors associated with CKD. By integrating these computational techniques, we aim to enhance the predictive accuracy of CKD onset and progression, thereby facilitating earlier intervention and better clinical outcomes.

The methodology of our research involves the application of four machine learning models: support vector machines (SVM), extreme gradient boosting (XGBoost), k-nearest neighbors (k-NN), and a stacking model. Each of these models brings a unique strength in handling complex, nonlinear relationships within large datasets, which is critical in accurately modeling the multifaceted nature of CKD risk factors. Our methodology includes a rigorous preprocessing phase involving imputation of missing values, one-hot encoding of categorical features, and removal of zero-variance features. Following this, we implemented and optimized three base learners: SVN, XGBoost, and k-NN. Further, we implemented GLMNet, random forest (RF), and XGBoost, and combined their outputs using a logistic regression-based stacking model. SVM has been extensively applied in medical diagnostics due to its robustness in dealing with high-dimensional data and its effectiveness in binary classification problems [11], [12]. For instance, Polat *et al.* [13] demonstrated the use of SVM to accurately classify stages of kidney disease using clinical datasets, highlighting its superior performance over other machine learning models in terms of precision and recall. Another study by Xiao *et al.* [14] applied SVM to differentiate between CKD patients and healthy controls with a high degree of accuracy, using genetic and laboratory data. Similarly, Singh *et al.* [15] employed SVM to explore the relationship between CKD and various demographic and biochemical markers, resulting in a model that could potentially guide early screening efforts in clinical settings. Additionally, Chen *et al.* [16] utilized SVM to analyze ultrasound images of the kidney, successfully identifying morphological changes associated with early stages of CKD. XGBoost has gained popularity for its ability to handle various types of data and its efficiency in performance on large datasets [17]. A study by Raihan *et al.* [18] employed XGBoost to identify CKD risk factors from electronic health records, achieving significant improvements in prediction accuracy compared to traditional models. Chuah *et al.* [19] applied XGBoost in a multi-center study to forecast renal function decline in CKD patients, where the model outperformed conventional risk scoring systems. Moreover, Minato *et al.* [20] demonstrated the effectiveness of XGBoost in predicting the outcomes of renal transplantation, an essential aspect of managing advanced CKD. Meanwhile, the k-NN algorithm is valued for its simplicity and effectiveness, particularly in scenarios where the relationship between variables is non-linear [21]. Devika *et al.* [22] employed k-NN to classify CKD severity based on biochemical and physiological parameters with high accuracy, demonstrating the effectiveness of this machine learning technique in identifying the condition's stages. Furthermore, Mahboob *et al.* [23] presented a framework for imputing missing values in large CKD datasets using *k*-Nearest Neighbors (k-NN), *k*-Means, and *k*-Medoids clustering algorithms, demonstrating that k-NN provided the most accurate results, achieving an accuracy of 86.67% with Decision Tree and 75.25% with RF, compared to the other algorithms. In addition to these models, we applied a stacking model, which combines the predictions of SVM, XGBoost, and k-NN using a logistic regression meta-learner to enhance predictive performance. A recent study by Mahajan *et al.* [24] have highlighted the potential of ensemble learning techniques in medical diagnostics, yet their application to CKD prediction remains underexplored. Furthermore, conflicting findings in the literature regarding the efficacy of different machine learning models underscore the need for comprehensive comparative analyses like ours.

Our study not only applies these models but also critically assesses their performance in the context of CKD prediction. We utilize a comprehensive dataset comprising clinical and demographic variables collected from patients suspected of having CKD. These include, but are not limited to, markers of kidney function such as serum creatinine levels, urea, and hemoglobin levels, as well as demographic factors like age and appetite. Our analysis aims to delineate the specific contributions of each factor to the risk of CKD, offering insights into both the biological and social determinants of kidney health. Furthermore, we aim to bridge the technological gap in public health applications by demonstrating the effectiveness of machine learning in improving CKD diagnosis and management, particularly in resource-limited settings. This

research has significant implications for early CKD diagnosis, particularly in settings with limited healthcare resources. By improving predictive accuracy, our approach can facilitate timely interventions and personalized treatment strategies, ultimately improving patient outcomes and reducing the burden on healthcare systems.

This paper is structured as follows: the Method section provides a detailed step-by-step description of the experimental procedure, including preprocessing steps, model implementation, and hyperparameter tuning. The Results and Discussion section presents the findings of our study, compares them with previous research, and discusses the implications and limitations of our work. Finally, the Conclusion summarizes our key contributions and suggests directions for future research.

2. METHOD

2.1. Dataset and software

In this study, we used the risk factor prediction of chronic kidney disease dataset from the UCI machine learning repository. The dataset consists of 28 features (excluding the class attribute) and contains 200 instances with a total of 5,600 patient data points. Table 1 shows the attributes of the dataset. The R programming software version 4.2.3 was used to do all the analysis in this paper.

Table 1. Dataset summary

Attribute	Description	Attribute	Description
Blood pressure (diastolic)	Diastolic blood pressure in mm Hg	Hemoglobin	Hemoglobin level in g/dL
Blood pressure limit	Blood pressure limit (categorical)	Packed cell volume	Packed cell volume (percentage)
Specific gravity	Urine specific gravity, ranging from 1.005 to 1.025	Red blood cell count	Red blood cell count in millions per cubic mm
Albumin	Protein levels in urine	White blood cell count	White blood cell count in cells per cubic mm
Red blood cells	Presence of red blood cells in urine	Hypertension	Presence of hypertension
Sugar	Sugar levels in urine	Diabetes mellitus	Presence of diabetes mellitus
Pus cell	Presence of pus cells in urine	Coronary artery disease	Presence of coronary artery disease
Pus cell clumps	Presence of pus cell clumps in urine	Appetite	Patient's appetite
Bacteria	Presence of bacteria in urine	Pedal edema	Presence of pedal edema
Blood glucose Random	Random blood glucose level in mg/dL	Anemia	Presence of anemia
Blood urea	Blood urea level in mg/dL	Glomerular filtration Rate	Estimated glomerular filtration rate (eGFR)
Sodium	Sodium level in blood in mEq/L	Stage	CKD stage (ranges from s1 to s5)
Serum creatinine	Serum creatinine level in mg/dL	Affected	Binary attribute indicating if the patient is affected by CKD
Potassium	Potassium level in blood in mEq/L	Age	Age of the patient in years

2.2. Support vector machine

SVM is a supervised machine learning model primarily used for classification tasks. It aims to find an optimal hyperplane that separates data points of different classes with the maximum margin. When data is not linearly separable, it leverages a kernel trick to map data into a higher-dimensional space. The Radial Basis Function kernel is particularly effective in capturing complex patterns [25].

An SVM model with an RBF kernel was developed using the e1071 and caret packages. To ensure robustness and reproducibility, we followed a meticulous preprocessing phase, which included imputing missing values, one-hot encoding categorical variables, and removing zero-variance features. We then split the dataset into training and testing sets, ensuring consistent class proportions through stratified sampling. The training set comprised 80% of the data, while the remaining 20% formed the test set. Numerical features were scaled to ensure consistent performance across different models.

Hyperparameters, including σ (kernel width) and C (regularization), are tuned using a grid search. The search will encompass a range of σ values (0.001, 0.01, 0.1, 0.5) and C values (2^2 to 2^9). The trainControl function is used for 10-fold cross-validation, and the train function from the caret package is used to build the model. Mathematically, SVM solves the following optimization problem:

$$\min_{w,b} \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\xi_i \quad (1)$$

subject to

$$y_i(\mathbf{w} \cdot \phi(\mathbf{x}_i) + b) \geq 1 - \xi_i, \ \ \xi_i \geq 0$$

where $\mathbf{w}$ is the weight vector, b is the bias term, $\phi(\mathbf{x}_i)$ is the non-linear mapping to a higher-dimensional space, C is the regularization parameter, and ξ_i are slack variables.

To further validate the model, we computed performance metrics such as accuracy, sensitivity, specificity, precision, recall, F1-Score, and ROC AUC. Moreover, we employed the permutation feature importance technique to identify the most influential features contributing to the model's predictions. This approach provided insights into the underlying data patterns and validated the importance of key clinical variables in predicting CKD.

2.3. Extreme gradient boosting

XGBoost is a scalable machine learning system for tree boosting. It implements a gradient-boosted decision tree algorithm with superior speed and performance. XGBoost sequentially builds a series of decision trees where each new tree attempts to correct errors made by the previous trees, resulting in a robust and accurate predictive model [26].

An XGBoost model was developed using the XGBoost package. To ensure the model's robustness and reproducibility, we performed meticulous preprocessing steps, including imputing missing values, one-hot encoding categorical variables, and removing zero-variance features. We then split the dataset into training and testing sets, maintaining consistent class proportions through stratified sampling. The training set comprised 80% of the data, while the remaining 20% formed the test set. Numerical features were scaled to enhance the model's performance. A grid search was conducted over various combinations of parameters such as nrounds, eta, max_depth, gamma, colsample_bytree, min_child_weight, and subsample. The trainControl function was used for 10-fold cross-validation, and the train function from the caret package was used to build the model. The objective function of XGBoost can be defined as

$$L(\mathbf{w}) = \sum_{i=1}^{n} l(\hat{y}_i, y_i) + \sum_{k=1}^{K} \Omega(f_k) \quad (2)$$

where l is the loss function, $\hat{y}_i$ is the predicted value, y_i is the true value, Ω is the regularization term, and f_k is the individual tree model. The model's performance was then evaluated on the test dataset and performance metrics were computed. Additionally, we conducted a feature importance analysis to identify the most influential features contributing to the model's predictions. This was achieved by plotting the feature importance, which highlighted the key variables driving the model's accuracy.

2.4. K-nearest neighbors

k-Nearest Neighbors is a non-parametric, instance-based learning method used for classification and regression. The algorithm classifies a data point based on the majority vote of the k nearest neighbors. The choice of k significantly affects the model's performance, with a lower value leading to more noise sensitivity and a higher value potentially diluting the decision boundaries [27].

The k-NN model was developed using the caret package. To ensure the robustness and reproducibility of our model, we undertook comprehensive preprocessing steps. This included imputing missing values, one-hot encoding categorical variables, and removing zero-variance features. The dataset was split into training and testing sets, maintaining consistent class proportions through stratified sampling. The training set comprised 80% of the data, while the remaining 20% formed the test set. Numerical features were scaled to improve model performance.

A grid search was conducted over the k parameter with odd values ranging from 1 to 30. The trainControl function was used for 10-fold cross-validation, and the train function from the caret package was used to build the model. k-NN classifies a new data point based on the majority vote of the k nearest neighbors:

$$\hat{y} = \arg\max_{y} \sum_{i \in N_k} 1(y_i = y) \quad (3)$$

where N_k is the set of k nearest neighbors and 1 is the indicator function. To validate the model, we computed the performance metrics. Additionally, we conducted a feature importance analysis to identify the most influential features contributing to the model's predictions.

2.5. Stacking model

Stacking is an ensemble learning technique that combines multiple classification models (base learners) to improve predictive performance by leveraging the strengths of each model [28], [29]. In this study, we implemented a stacking model using three base learners: generalized linear model (GLMNet), RF,

and XGBoost. These base learners were chosen for their diverse strengths in handling different data patterns and complexities.

To ensure the robustness and reproducibility of our model, we undertook comprehensive preprocessing steps. This included imputing missing values, one-hot encoding categorical variables, and removing zero-variance features. The dataset was then split into training and testing sets, maintaining consistent class proportions through stratified sampling. The training set comprised 80% of the data, while the remaining 20% formed the test set. Numerical features were scaled to improve model performance. The base learners were trained using the train function from the caret package with 10-fold cross-validation to ensure generalizability. The hyperparameters for each base learner were tuned using grid search:

- GLMNet: Regularized logistic regression model tuned for optimal lambda and alpha values.
- RF: An ensemble of decision trees tuned for the number of trees and maximum features.
- XGBoost: Gradient boosting algorithm tuned for nrounds, eta, max_depth, gamma, colsample_bytree, min_child_weight, and subsample.

After training the base learners, predictions were made on the test dataset. These predictions were then combined into a new dataset, which served as input for the meta-learner. The meta-learner was trained using logistic regression with 10-fold cross-validation, using the combined predictions from the base learners to make the final classification. This approach allowed the meta-learner to learn from the strengths and weaknesses of each base learner, resulting in a more robust and accurate predictive model. Figure 1 presents the diagram of the proposed stacking model.

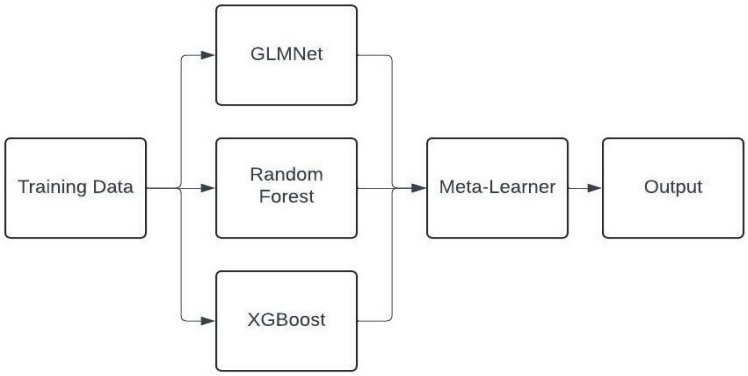


Figure 1. Proposed stacking model flowchart

2.6. Model performance metrics

The performance of each model was assessed using several evaluation metrics to provide a comprehensive analysis of their predictive capabilities. The following metrics were employed to evaluate the effectiveness and reliability of the models:

Accuracy measures the overall correctness of the model and is calculated as:

$$\text{Accuracy}: \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

Precision (positive predictive value) indicates the proportion of positive predictions that are actually correct:

$$\text{Precision (pos pred value)}: \frac{TP}{TP+FP} \quad (5)$$

Recall (sensitivity) reflects the proportion of actual positives that are correctly identified by the model:

$$\text{Recall (sensitivity)}: \frac{TP}{TP+FN} \quad (6)$$

F1 Score provides a harmonic mean of precision and recall, offering a balance between the two metrics:

$$\text{F1 Score}: \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

where TP are true positives, TN are true negatives, FP are false positives, and FN are false negatives. These metrics collectively offer a detailed view of the model's performance, addressing various aspects of prediction accuracy and error rates. Each metric provides unique insights into different performance aspects, thereby ensuring a comprehensive evaluation of the models.

3. RESULTS AND DISCUSSION

The initial dataset contained 200 samples and 28 features representing clinical and demographic attributes relevant to chronic kidney disease. Preprocessing was done to the data before implementing the models. After replacing placeholder strings with NA values and imputing missing numerical data using median imputation, the dataset was one-hot encoded to transform categorical features into numerical ones. Features with zero variance were removed, as they do not provide any useful information for the model and can lead to overfitting. This resulted in a final dataset of 161 samples and 143 features, which included one-hot encoded and numerical attributes. After preprocessing, the dataset was divided into training and testing sets using stratified sampling to ensure consistent class proportions. The training set contained 129 samples, while the testing set had 32 samples. Numerical features in the training and testing sets were scaled to ensure consistent performance across different machine learning models.

The SVM model was implemented using the svmRadial method in the caret package. Grid search was used to tune the hyperparameters σ and C. The best combination of parameters was found to be $\sigma = 0.01$ and $C = 4$. The SVM model achieved an accuracy of 100%, indicating that it correctly classified all test samples. The Kappa statistic of 1 signifies perfect agreement between predicted and actual classes. The high sensitivity (1.000) and specificity (1.000) demonstrate that the model can accurately distinguish between CKD and non-CKD samples. Additionally, the positive predictive value (1.000) and negative predictive value (1.000) confirm the model's reliability in identifying both CKD-positive and CKD-negative cases. Table 2 presents the evaluation metrics of the SVM model.

Table 2. Evaluation metrics of the SVM model

Metric	Value
Accuracy	1.0000
95% confidence interval	(0.9097,1)
No information rate	0.6410
Kappa	1.0000
Sensitivity	1.0000
Specificity	1.0000
Positive predictive value	1.0000
Negative predictive value	1.0000
Prevalence	0.6410
Detection rate	0.6410
Detection prevalence	0.6410
Balanced accuracy	1.0000

The XGBoost model was implemented using the xgbTree method in the caret package. Grid search and cross-validation were used to tune the following hyperparameters: nrounds, max_depth, eta, gamma, colsample_bytree, min_child_weight, and subsample. The best parameter combination is shown in Table 3.

Table 3. XGBoost best parameters

Parameter	Value
nround	50
max_depth	3
eta	0.01
gamma	0
colsample_bytree	0.5
min_child_weight	1
subsample	0.5

As presented in Table 4, the XGBoost model exhibited strong classification performance, achieving an accuracy of 97.44% with a kappa statistic of 0.9451, signifying excellent agreement between predicted and actual classifications. The model's sensitivity (0.9600) and specificity (1.0000) demonstrate its ability to distinguish between CKD and non-CKD samples effectively. The positive predictive value (1.0000) and negative predictive value (0.9333) confirm the model's reliability in identifying CKD-positive and CKD-negative cases.

Table 4. Evaluation metrics of the XGBoost model

Metric	Value
Accuracy	0.9744
95% confidence interval	(0.8652, 0.9994)
No information rate	0.6410
Kappa	0.9451
Sensitivity	0.9600
Specificity	1.0000
Positive predictive value	1.0000
Negative predictive value	1.0000
Prevalence	0.6410
Detection rate	0.6154
Detection prevalence	0.6154
Balanced accuracy	0.9800

The k-NN model was implemented using the k-NN method in the caret package. Grid search was used to find the optimal value for k. The best parameter found is $k = 21$. The k-NN model was then trained with the optimal k value, and predictions were made on the test data. Table 5 provides the performance metrics. The k-NN model achieved a strong overall performance, with an accuracy of 92.31% and a kappa statistic of 0.8404, indicating substantial agreement between predictions and actual labels. The model's sensitivity (0.8800) and specificity (1.0000) reflect its effectiveness in identifying CKD and non-CKD cases. The positive predictive value (1.0000) and negative predictive value (0.8235) highlight the model's ability to predict CKD accurately while minimizing false negatives. The balanced accuracy of 0.9400 further demonstrates the model's robust classification ability.

Table 5. Evaluation metrics of the k-NN model

Metric	Value
Accuracy	0.9744
95% confidence interval	(0.8652, 0.9994)
No information rate	0.6410
Kappa	0.9451
Sensitivity	0.9600
Specificity	1.0000
Positive predictive value	1.0000
Negative predictive value	1.0000
Prevalence	0.6410
Detection rate	0.6154
Detection prevalence	0.6154
Balanced accuracy	0.9800

The stacking model was implemented to leverage the strengths of multiple base learners: GLMNet, RF, and XGBoost. The base learners were trained using the train function from the caret package with 10-fold cross-validation. After training, predictions from the base learners were combined into a new dataset, which served as input for the meta-learner, a logistic regression model trained with 10-fold cross-validation.

The stacking model achieved an impressive performance on the test dataset, with an accuracy of 100%. The confusion matrix showed perfect agreement between predicted and actual classes, with a Kappa statistic of 1, indicating no difference between the observed agreement and perfect agreement. The high sensitivity (1.000) and specificity (1.000) demonstrate the model's effectiveness in correctly identifying CKD and non-CKD cases. Additionally, the precision (1.000), recall (1.000), F1-Score (1.000), and ROC AUC (1.000) metrics highlight the model's robustness and reliability. Table 6 presents the other evaluation metrics of the stacking model.

Table 6. Evaluation metrics of the stacking model

Metric	Value
Accuracy	1.0000
95% confidence interval	(0.9097, 1.0000)
No information rate	0.6410
Kappa	1.0000
Sensitivity	1.0000
Specificity	1.0000
Positive predictive value	1.0000
Negative predictive value	1.0000
Prevalence	0.6410
Detection rate	0.6410
Detection prevalence	0.6410
Balanced accuracy	1.0000

- Comparative Analysis

Table 7 summarizes the comparative analysis of the models used in this study. The SVM model achieved perfect scores across all metrics, while the XGBoost model demonstrated high precision and recall values, misclassifying only one sample. The k-NN model exhibited slightly lower recall compared to SVM and XGBoost, leading to a reduced F1-score. Finally, the Stacking model achieved perfect performance.

Figure 2 provides a visualization of feature importance for the XGBoost model. It is revealed that albumin (protein levels in urine), the presence of diabetes mellitus, urine specific gravity, and hypertension are significant risk factors for CKD. On the other hand, blood urea and random blood glucose are less significant risk factors for CKD. The features found in the XGBoost model were similar to those identified by the SVM, k-NN, and stacking models, Figure 1 is sufficient, and we will not show the results of feature importance for the other models.

Table 7. Comparative analysis of models

Attribute	SVM	XGBoost	k-NN	Stacking model
Precision	1.000	0.9744	0.9231	1.000
Recall	1.000	0.9600	0.8800	1.000
F1-score	1.000	0.9800	0.936	1.000

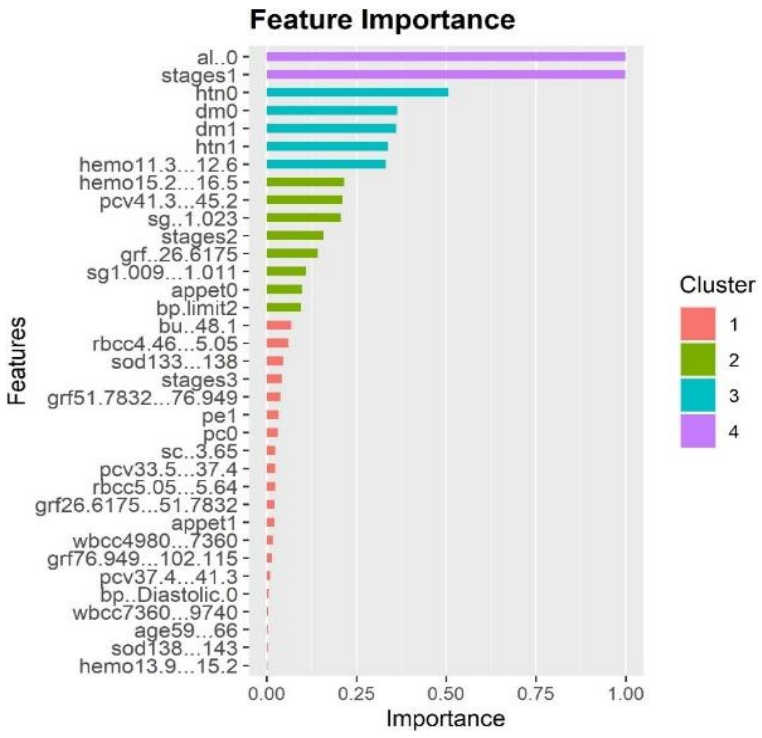


Figure 2. Feature importance from XGBoost model

The results demonstrate the efficacy of various machine learning models in predicting CKD. The SVM model achieved perfect classification performance, reflecting its robustness in handling high-dimensional data. The XGBoost model also performed excellently, showcasing its ability to handle complex data patterns and interactions. Although the k-NN model had slightly lower performance compared to SVM and XGBoost, it still showed strong predictive capabilities, especially in identifying CKD-positive cases.

The stacking model, which combined the predictions from GLMNet, RF, and XGBoost using a logistic regression meta-learner, also has a perfect performance. This highlights the strength of ensemble learning techniques in leveraging the strengths of multiple models to improve overall predictive performance. The perfect accuracy, sensitivity, and specificity achieved by the stacking model demonstrate its potential as a reliable tool for CKD prediction.

The high accuracy and reliability of the machine learning models, particularly the stacking model, suggest that these techniques can be effectively used in clinical settings to predict CKD onset. Early and accurate prediction of CKD can facilitate timely interventions, personalized treatment strategies, and ultimately improve patient outcomes. The integration of machine learning models in healthcare can also help in resource-limited settings by providing decision support to healthcare professionals.

Our findings are consistent with previous studies that have demonstrated the potential of machine learning in CKD prediction. For instance, Polat *et al.* [13] and Singh *et al.* [15] highlighted the superior performance of SVM in CKD classification. However, our SVM model performs better, achieving higher accuracy compared to their models, which reported accuracies of 97.75% and 92%, respectively. Similarly, Raihan *et al.* [18] and Chuah *et al.* [19] demonstrated the efficacy of XGBoost in predicting CKD and renal function decline, respectively. When compared to the performance of their models, the XGBoost model of Raihan *et al.* [18] performed better with an accuracy of 99.16% compared to our 97.44%. However, our model performed better than that of Chuah *et al.* [19], whose XGBoost model had an accuracy of 93.9%. Our study extends these findings by incorporating a stacking model that combines multiple base learners to achieve higher predictive accuracy. The efficacy of our stacking model is similar to the study by Bhagyalaxmi and Dwarakanath [30], which showed that combining the strengths of different models, such as in hybrid models, improves performance. In the study, a hybrid CNN-LR model outperformed other classifiers in brain tumor detection, highlighting the effectiveness of leveraging multiple algorithms. Similarly, our stacking model's superior performance underscores the advantage of integrating different machine learning techniques to enhance predictive capabilities. This superior performance is also similar to the performances of hybrid models found in the studies of Gupta and Sharma [31], who developed a hybrid CNN and LSTM model for heart disease prediction, and used a hybrid CNN and fuzzy kernel K-medoids model for lung cancer detection. These studies collectively demonstrate that models which combine various machine learning techniques, offer enhanced accuracy and reliability in medical diagnosis.

A key strength of this study is the rigorous preprocessing and comprehensive evaluation of multiple machine learning models, including the use of a stacking model. The high performance of the models emphasizes the importance of thorough data preprocessing and hyperparameter tuning. However, the study has some limitations. The dataset used was relatively small and may not fully capture the diversity of CKD patients. Additionally, the models were evaluated on a single dataset, and their generalizability to other populations needs to be tested.

Future research should incorporate larger, more diverse datasets and explore real-world clinical data to validate the models further. Additionally, investigating other advanced machine learning techniques could provide further insights. Integrating these models into clinical practice involves developing user-friendly decision support systems for healthcare professionals and conducting pilot studies in clinical settings to assess their impact on CKD diagnosis and management, ensuring practical applicability and effectiveness in real-world scenarios.

4. CONCLUSION

This study highlighted the efficacy of SVM, Extreme Gradient Boosting, *k*-Nearest Neighbors, and a Stacking Model in predicting the onset of chronic kidney disease. The SVM model achieved perfect classification performance with 100% accuracy, precision, and recall, while XGBoost closely followed with an accuracy of 97.44% and a kappa of 0.9451. The k-NN model also delivered strong performance with an accuracy of 92.31%. The stacking model, which leveraged the strengths of multiple base learners, outperformed all individual models, achieving perfect accuracy and demonstrating its potential as a reliable tool for CKD prediction. These results emphasize the significant potential of machine learning models in healthcare, particularly for early detection and diagnosis of CKD. Accurate and timely prediction can lead to improved clinical outcomes by enabling earlier interventions and personalized treatment strategies. The integration of machine learning models in healthcare can also help in resource-limited settings by providing

decision support to healthcare professionals. The success of these models in handling the complex, nonlinear relationships inherent in CKD datasets suggests that further integration of machine learning into nephrology could substantially enhance predictive capabilities. However, the study has some limitations, such as the relatively small dataset and its limited diversity, which may affect the generalizability of the findings. Future research should emphasize the inclusion of more comprehensive clinical features, improved data collection methods, and feature engineering to bolster predictive accuracy. Additionally, expanding model training to larger and more diverse datasets would improve the generalizability of these findings, ultimately aiding in the global fight against CKD and improving patient outcomes. By addressing these future research directions, the practical applicability and impact of machine learning models in nephrology can be significantly enhanced, contributing to better health outcomes for CKD patients globally.

REFERENCES

[1] A. Sanyaolu *et al.*, "Epidemiology and management of chronic renal failure: a global public health problem," *Biostatistics and Epidemiology International Journal*, vol. 1, no. 1, pp. 11–16, 2018.

[2] J. C. Lv and L. X. Zhang, "Prevalence and disease burden of chronic kidney disease," *Advances in Experimental Medicine and Biology*, vol. 1165, pp. 3–15, 2019, doi: 10.1007/978-981-13-8871-2_1.

[3] A. Francis *et al.*, "Chronic kidney disease and the global public health agenda: an international consensus," *Nature Reviews Nephrology*, vol. 20, no. 7, pp. 473–485, Jul. 2024, doi: 10.1038/s41581-024-00820-6.

[4] D. Rucker *et al.*, "Quality of care and mortality are worse in chronic kidney disease patients living in remote areas," *Kidney International*, vol. 79, no. 2, pp. 210–217, 2011, doi: 10.1038/ki.2010.376.

[5] F. F. Kaze, D. T. Meto, M. P. Halle, J. Ngogang, and A. P. Kengne, "Prevalence and determinants of chronic kidney disease in rural and urban Cameroonians: A cross-sectional study," *BMC Nephrology*, vol. 16, no. 1, 2015, doi: 10.1186/s12882-015-0111-8.

[6] E. S. Wijewickrama *et al.*, "Prevalence of CKD of unknown etiology and its potential risk factors in a rural population in Sri Lanka," *Kidney International Reports*, vol. 7, no. 10, pp. 2303–2307, 2022, doi: 10.1016/j.ekir.2022.07.012.

[7] S. Chintala, "The role of ai in predicting and managing chronic diseases," *International Journal of New Media Studies (IJNMS)*, vol. 7, no. 2, pp. 16–22, 2020.

[8] K. Singhania and A. Reddy, "Improving preventative care and health outcomes for patients with chronic diseases using big data-driven insights and predictive modeling," *International Journal of Applied Health Care Analytics*, vol. 9, no. 2, pp. 1–14, 2024, [Online]. Available: https://norislab.com/index.php/IJAHA/article/view/60

[9] L. K. Pathak and P. Jha, "Application of machine learning in chronic kidney disease risk prediction using electronic health records (EHR)," in *In Applications of Big Data in Large-and Small-Scale Systems*, IGI Global Scientific Publishing, 2021, pp. 213–233, doi: 10.4018/978-1-7998-6673-2.ch014.

[10] M. M. Ahsan, S. A. Luna, and Z. Siddique, "Machine-learning-based disease diagnosis: a comprehensive review," *Healthcare (Switzerland)*, vol. 10, no. 3, p. 541, 2022, doi: 10.3390/healthcare10030541.

[11] R. Aziz, C. K. Verma, and N. Srivastava, "Artificial neural network classification of high dimensional data with novel optimization approach of dimension reduction," *Annals of Data Science*, vol. 5, no. 4, pp. 615–635, 2018, doi: 10.1007/s40745-018-0155-2.

[12] A. Kalantari, A. Kamsin, S. Shamshirband, A. Gani, H. Alinejad-Rokny, and A. T. Chronopoulos, "Computational intelligence approaches for classification of medical data: State-of-the-art, future challenges and research directions," *Neurocomputing*, vol. 276, pp. 2–22, 2018, doi: 10.1016/j.neucom.2017.01.126.

[13] H. Polat, H. D. Mehr, and A. Cetin, "Diagnosis of chronic kidney disease based on support vector machine by feature selection methods," *Journal of Medical Systems*, vol. 41, no. 4, pp. 1–11, 2017, doi: 10.1007/s10916-017-0703-x.

[14] J. Xiao *et al.*, "Comparison and development of machine learning tools in the prediction of chronic kidney disease progression," *Journal of Translational Medicine*, vol. 17, no. 1, p. 119, 2019, doi: 10.1186/s12967-019-1860-0.

[15] V. Singh, V. K. Asari, and R. Rajasekaran, "A deep neural network for early detection and prediction of chronic kidney disease," *Diagnostics*, vol. 12, no. 1, p. 116, 2022, doi: 10.3390/diagnostics12010116.

[16] C. J. Chen, T. W. Pai, H. H. Hsu, C. H. Lee, K. S. Chen, and Y. C. Chen, "Prediction of chronic kidney disease stages by renal ultrasound imaging," *Enterprise Information Systems*, vol. 14, no. 2, pp. 178–195, 2020, doi: 10.1080/17517575.2019.1597386.

[17] A. Asselman, M. Khaldi, and S. Aammou, "Enhancing the prediction of student performance based on the machine learning XGBoost algorithm," *Interactive Learning Environments*, vol. 31, no. 6, pp. 3360–3379, 2023, doi: 10.1080/10494820.2021.1928235.

[18] M. J. Raihan, M. A. M. Khan, S. H. Kee, and A. Al Nahid, "Detection of the chronic kidney disease using XGBoost classifier and explaining the influence of the attributes on the model using SHAP," *Scientific Reports*, vol. 13, no. 1, p. 6263, 2023, doi: 10.1038/s41598-023-33525-0.

[19] A. Chuah *et al.*, "Machine learning improves upon clinicians' prediction of end stage kidney disease," *Frontiers in Medicine*, vol. 9, p. 837232, 2022, doi: 10.3389/fmed.2022.837232.

[20] A. C. dos S. Minato *et al.*, "Machine learning model to predict graft rejection after kidney transplantation," in *Transplantation Proceedings*, 2023, vol. 55, no. 9, pp. 2058–2062, doi: 10.1016/j.transproceed.2023.07.021.

[21] S. Zhang, D. Cheng, Z. Deng, M. Zong, and X. Deng, "A novel kNN algorithm with data-driven k parameter computation," *Pattern Recognition Letters*, vol. 109, pp. 44–54, 2018, doi: 10.1016/j.patrec.2017.09.036.

[22] R. Devika, S. V. Avilala, and V. Subramaniyaswamy, "Comparative study of classifier for chronic kidney disease prediction using naive bayes, KNN and random forest," in *Proceedings of the 3rd International Conference on Computing Methodologies and Communication, ICCMC 2019*, 2019, pp. 679–684, doi: 10.1109/ICCMC.2019.8819654.

[23] T. Mahboob, A. Ijaz, A. Shahzad, and M. Kalsoom, "Handling missing values in chronic kidney disease datasets using KNN, k-means and k-medoids algorithms," in *ICOSST 2018 - 2018 International Conference on Open Source Systems and Technologies, Proceedings*, 2018, pp. 76–81, doi: 10.1109/ICOSST.2018.8632179.

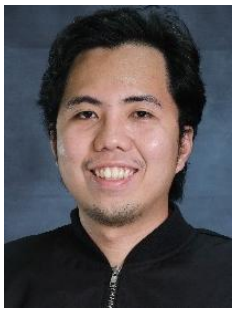
[24] P. Mahajan, S. Uddin, F. Hajati, and M. A. Moni, "Ensemble learning for disease prediction: a review," *Healthcare (Switzerland)*, vol. 11, no. 12, p. 1808, 2023, doi: 10.3390/healthcare11121808.

[25] C.-X. Deng, L.-X. Xu, and S. Li, "Classification of support vector machine and regression algorithm," *New Advances in Machine Learning*, 2010, doi: 10.5772/9392.

[26] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785–794, doi: 10.1145/2939672.2939785.

[27] M. Cover T and E. Hart P, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, 1967.
[28] O. Sagi and L. Rokach, "Ensemble learning: A survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 8, no. 4, p. e1249, 2018, doi: 10.1002/widm.1249.
[29] P. Mahajan, S. Uddin, F. Hajati, and M. A. Moni, "Ensemble learning for disease prediction: a review," *Healthcare (Switzerland)*, vol. 11, no. 12, p. 1808, 2023, doi: 10.3390/healthcare11121808.
[30] K. Bhagyalaxmi and B. Dwarakanath, "Hybrid model for detection of brain tumor using convolution neural networks," *Computer Science and Information Technologies*, vol. 5, no. 1, pp. 84–90, 2024, doi: 10.11591/csit.v5i1.pp78-84.
[31] S. Gupta and P. Sharma, "Hybrid model of convolutional neural network and long short term memory for heart disease prediction," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 33, no. 1, pp. 389–397, 2024, doi: 10.11591/ijeecs.v33.i1.pp389-397.

BIOGRAPHIES OF AUTHORS



Samuel John Parreño is a faculty member at the University of Mindanao Digos College, Digos City, Philippines. He is a Licensed Professional Teacher and placed first in the Licensure Examination for Teachers in 2018. He received his Bachelor's Degree in Secondary Education, majoring in mathematics, from the University of Mindanao Digos College and an M.Sc. degree in Applied Mathematics from the University of Southeastern Philippines. He is a member of learned societies, including the Society for Industrial and Applied Mathematics (Philadelphia, USA), the Southeast Asian Mathematical Society, and the Mathematical Society of the Philippines. His research area is in applied mathematics, covering computational topology, data science, machine learning, mathematical statistics, and computational biology, with particular interest in the application of mathematics in biology, epidemiology, and pharmacometrics. He can be contacted at email: samueljohn_parreno@umindanao.edu.ph.

Maria Cristine Joy Anter is a faculty member of the Teacher Education Department at University of Mindanao Digos College and a Licensed Professional Teacher. She holds a Bachelor's degree in Secondary Education from University of Mindanao Digos College and a Master of Arts in Teaching Mathematics from the University of Southeastern Philippines. She is a member of the Mathematics Teachers Association of the Philippines Tertiary Level (MTAPTL). Her research focuses on education, mathematics, data science, machine learning, and mathematical statistics. Her work aims to bridge traditional educational methodologies with modern technological advances. She can be contacted at email: mariacristinejoyanter@umindanao.edu.ph.