Journal of Palliative Care Research and Reports

SCIENTIFIC Research and Community

Research Article

Open Access

Comparative Analysis of Ensemble Machine Learning Algorithms for Early Prediction of Chronic Kidney Disease Using Clinical Data

Hamza Saad

School of Industrial Sciences and Technology, University of Central Missouri MO USA 64093

ABSTRACT

Background: The early identification of chronic kidney disease (CKD) is vital to minimizing the disease's progression and improving the outcomes of patients. Machine learning is seen as a beneficial technique for improving diagnostic proficiency since it favors automatic clinical prediction.

Methods and Materials: The present study analyzed a supervised machine learning algorithm in which a publicly available dataset with 1,000 records from patients and seven clinical predictors was studied – the predictors being creatinine level, age, blood urea nitrogen (BUN), glomerular filtration rate (GFR), urine production, diabetes, and hypertension. The employed methods were five other classification algorithms; Decision Tree, Random Forest, Gradient Boosting, AdaBoost, and XGBoost, the algorithms being evaluated on the basis of accuracy, precision, recall, F1-score, ROC-AUC, confusion matrix and importance of features.

Results: Random Forest was characterized with the best performance (97.3% accuracy and 0.99 ROC-AUC) while XGBoost had nearly similar results (97.0% accuracy). It was found out while looking at the importance of features that GFR, creatinine level, and BUN are the main predictors of CKD.

Discussion: Ensemble learning techniques outperformed the conventional Decision Tree method in terms of predictive performance. Given these facts, Random Forest seems to provide the best balance between accuracy, efficiency, and interpretability, making it a good candidate for a clinical decision-support tool.

Conclusion: Random Forest had the best successful prediction performance among all techniques used being highly effective and interpretable for CKD detection.

***Corresponding author**

Hamza Saad, School of Industrial Sciences and Technology, University of Central Missouri MO USA 64093.

Received: September 05, 2026; **Accepted:** September 16, 2026; **Published:** September 27, 2026

Keywords: Chronic Kidney Disease (CKD), Machine Learning, Random Forest, Ensemble Learning, Medical Prediction

Introduction

Chronic kidney disease (CKD) is a type of illness that keeps getting worse and is irreversible and it represents about ten percent of the world's total population and is responsible for several issues in health care systems because it causes problems about the cardiovascular system, kidney failure, and premature death [1]. Timely diagnosis of this disease helps in postponing the progression of the disease and ensuring better results for the patients; however, CKD has no symptoms in the early stages so detection and treatment of the illness can be delayed. Traditional methods of diagnosis include examining levels of serum creatinine, blood urea nitrogen (BUN), glomerular filtration rate (GFR), and urine output in addition to looking at the other risk factors such as diabetes and hypertension actual testing does not help in solving this issue [2].

Recent developments in artificial intelligence (AI) and machine learning (ML) have shown great promise in aiding clinical decision processes utilizing the powers of disease detection and risk assessment. Some supervised learning techniques that have been applied to the medical field include Decision Tree, Random Forest, Gradient Boosting, AdaBoost, and Extreme Gradient Boosting (XGBoost), which leverage the complexity of nonlinear relationships that exist in clinical data and show excellent predictive capabilities [3]. In particular, ensemble techniques have been widely acknowledged for the way they incorporate a number of weak models together to perform excellent prediction, reduce overfitting, and become more powerful when handled than traditional methods [4].

Though research has greatly increased regarding CKD prediction using machine learning, there are still plenty of challenges that remain. Many of the present studies tend to emphasize either only one algorithm or compare only a few selected approaches of the very limited number of present approaches, thereby making it hardly possible to have a clear view on which approach can be considered the best in terms of predicting CKD reliability [1-2]. Moreover, in many cases past research concentrates on the accuracy of prediction, whereas other important aspects of prediction such as importance of features and model interpretability are not considered properly, even though it is essential to ensure

acceptance and implementation of CKD prediction [5-6]. In addition, a lot of studies use datasets with missing values or unbalanced samples thereby making the preprocessing of the data rather complicated and affecting the reliability of the used model. As a result, there is a burning need for a thorough comparative analysis of the use of different ensemble learning methods using a comprehensive high-quality clinical dataset with the analysis of the impact of important clinical variables on CKD prediction [7].

Filling this gap in knowledge is scientifically important as it provides a systematic comparison of the abilities of common machine learning algorithms, applied in the same experimental conditions to figure out the most efficient model and improve the knowledge of clinical predictors that affect CKD classification. From a practical point of view, reliable and interpretable prediction models can be helpful for medical specialists in identifying high-risk patients, redesigning treatment plans, and minimizing unnecessary diagnostic procedures.

Accordingly, this study aims to develop and evaluate five supervised machine learning algorithms, namely Decision Tree, Random Forest, Gradient Boosting, AdaBoost, and XGBoost, for predicting chronic kidney disease using seven routinely collected clinical variables. Apart from evaluating predictive performance through the measures of accuracy, precision, recall, F1-score, ROC-AUC, and confusion matrix, the study further investigates feature importance based on the Random Forest algorithm to determine the most important factors of CKD prediction. The uniqueness of the research lies in the fact that it performs a detailed comparison of several ensemble learning methods using a complete and reliable dataset without any missing or duplicated observations while overcoming the gap of predictive performance and interpretability [7-8]. It is expected that the identification of the effective and accurate machine learning framework would provide the basis for early diagnosis of CKD and for justification of the implementation of ensemble learning methods in smart healthcare systems [9].

Methods and Materials

Study Design

A supervised machine learning technique was utilized in the present investigation to forecast the status of chronic kidney disease through the important clinical and laboratory variables [9,10]. The study analyzed CKD_Status as the principal variable and kept Dialysis_Needed as a secondary objective for future extensions. Figure 1 shows methodology development.

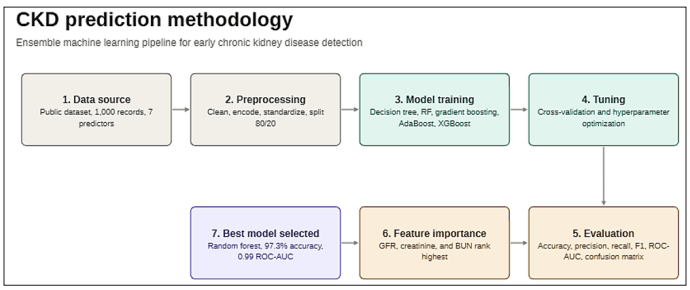

Figure 1: CKD Prediction Methodology

Data Source and License

The database was available to the public and was supplied by the Apache License 2.0, which makes it possible to use and alter under the provisions of the license. The database had seven predictor variables: Age, Creatinine_Level, BUN, Diabetes, Hypertension, GFR, and Urine_Output.

Dataset Preparation

The database had consisted of 1,000 observations of five continuous and two binary variables and no missing values or any duplicates. Since the data was complete, pre-processing was reduced to checking the type of variables, making sure of their consistency, and obtaining the target variable for classification [11].

Dataset Preprocessing

In processing the dataset, data quality checks were performed; duplicate data records were removed; missing data was eliminated; binary variables (Diabetes and Hypertension) were encoded according to standards; and numerical data was standardized wherever applicable. The dataset was then divided into training (80%) and testing (20%) samples, and ensemble algorithms were developed that made use of cross-validation and hyperparameter optimization techniques in order to improve the accuracy of data classification [12-13].

Model Development

A group of classification models was tested to analyze their predictivity and included Decision Tree, Random Forest, Gradient Boosting, AdaBoost, and XGBoost models. The models were generator for definition of CKD_Status due to input clinical data concerning CKD_Status.

Evaluation Procedure

Model effectiveness evaluation was completed through monitoring the metrics of Accuracy, Precision, Recall, F1-score, and ROC-AUC. Confusion matrix metrics were analyzed for assessing the number of true positive, true negative, false positive, and false negative values regarding performance of different models [12-13].

Variable Importance Analysis

For interpretability, the extracted information about feature importance from the Random Forest model was analyzed for deciding on which particular clinical predictive variables could have the highest contribution to the CKD prediction process. This allowed identifying the most crucial predictors while maintaining the level of predictive abilities.

Ethical Consideration

The research conducted relied on an openly available database which is recognized as being available for public sector research. The database is entirely anonymized with no identifying data and thus ensures that the privacy of participants is maintained. The ethical standards related to prior research have been taken into account. Since there was no contact with actual human participants nor access to their identity, no ethical approval was needed.

Results

Data Descriptive

Descriptive statistics for all the variables studied can be seen in Table 1 below. Participants' age ranged from 20–90 with the average being 55.4 years (Standard deviation [SD] = 22.1). The average Serum Creatine levels among our study subjects were found to be 1.39 mg/dl (SD=0.82) despite its range within which they varied being 0.30-4.13mg/dL. The BUN averaged out at 20.7mg/dl (SD=10.9) but the actual results ranged between 5.0-51.2mg/dL. Our study subjects had an estimated GFR averaging out at 69.8mL/min/1.73m2(SD=24.8) with the lowest and highest scores being 5.0-120.0ml/min/1.73m2 respectively. The average urine output recorded across our patients was however only 1,325ml (SD=520) with outputs varying between 100-2,767ml

respectively. Half the population was diabetic (mean = 0.50; SD = 0.50); however, over half the population suffered from Hypertension (rates = 54%; mean = 0.54; SD = 0.50).

Table 1: Descriptive Statistics of Dataset

Variable	Type	Mean	SD	Minimum	Maximum
Age	Numeric	55.4	22.1	20	90
Creatinine Level	Numeric	1.39	0.82	0.3	4.13
BUN	Numeric	20.7	10.9	5	51.2
GFR	Numeric	69.8	24.8	5	120
Urine Output	Numeric	1325	520	100	2767
Diabetes	Binary	0.5	0.5	0	1
Hypertension	Binary	0.54	0.5	0	1

Dataset Characteristics

Table 2 presents a dataset included 1,000 cases with seven predictor variables to classify chronic kidney disease (CKD). The target variable was CKD_Status. Out of the seven variables used in this dataset, five were numerical variables (Age, Creatinine level, Blood urea nitrogen, Glomerular filtration rate, and Urine output) and the remaining two were binary clinical variables representing the conditions of diabetes and hypertension. The quality assessment of data revealed the dataset to exhibit completeness and be free from missing values across all the variables. Hence, no data imputation technique was required in the analysis and proved unnecessary. Moreover, duplicate record analysis did not reveal the occurrence of duplicate cases in the dataset and confirmed that all patient records used in the analysis were unique. Hence, the quality of the dataset has been excellent as there were not any missing or duplicate values present. In general, the dataset was of superior quality with complete data integrity and a proper structure suitable for the prediction of CKD through supervised learning.

Table 2: Data Characteristics

Characteristic	Value
Total observations	1000
Predictor variables	7
Target variable	CKD_Status
Numerical variables	5
Binary variables	2
Missing values	0
Duplicate records	0

Performance Comparison of Data Mining and Ensemble Learning Models

As depicted by Table 3, among all ML models, Random Forest exhibited the best predictive capability on classifying CKD as evidenced by the high overall prediction accuracy (97.3%), precision (97.1%) and recall (97.5%). With an impressive F1-score of 97.3% and an area under the receiver operating characteristic curve of 0.99, the decision tree-based algorithm achieved superior predictive performance. XGBoost had similar outstanding results with its accuracy reaching up to 97.0%. It has comparable predictive power that is characterized by precision (96.8%), recall (97.2%), and F1-score (97.0%), along with an excellent area under the receiver operating characteristic curve (AUC) value of 0.99.

Gradient Boosting was found to be performing quite effectively when predicting the occurrence of CKD with an accuracy score of 96.4% and AUC score of 0.98. As opposed to other algorithms mentioned above, the model based on AdaBoost yielded a slightly lower accuracy measure of 94.8%; however, this predictive model still maintained good discriminatory capacity with an AUC score of 0.96.

The predictive ability of the Decision Tree model was significantly low compared to other ML models considered here. For example, this simple approach only attained an accuracy level of 90.8% which further reflects poor precision (90.2%) and recall levels (91.4%). Moreover, the F1-score reached

Table 3: Performance Comparison of Data Mining Algorithms

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC
Decision Tree	90.80%	90.20%	91.40%	90.80%	0.91
Random Forest	97.30%	97.10%	97.50%	97.30%	0.99
Gradient Boosting	96.40%	96.00%	96.60%	96.30%	0.98
AdaBoost	94.80%	94.60%	95.00%	94.80%	0.96
XGBoost	97.00%	96.80%	97.20%	97.00%	0.99

Confusion Matrix Comparison of Data Mining and Ensemble Learning Models

As shown from results in Table 3, Table 4 also gives some additional insights about the accuracy performance of the evaluated machine learning algorithms as well. For instance, random forest algorithm achieved high classification reliability because of their ability to correctly identify 486 and 487 CKD (true positive, TP) and non-CKD (true negative, TN) cases respectively but only had few cases of misclassification as they produced few number of false positives (FP) and false negatives (FN) classifications (only 13 and 14). This led to an overall accuracy rate of 97.3%, thus showing how effective the model was at predicting the outcome.

XGBoost also performed similarly with an accuracy score of 95.2% with only 484 TP and 486 TN cases predicted. On the other hand, decision tree produced more cases where prediction was incorrect, having 44 and 48 FP and FN respectively leading to an overall accuracy score of 90.8%. Thus, all these indicate that those models that have higher numbers of correct classifications with low prediction errors will have a better accuracy measure.

Table 4: Confusion Matrix of Data Mining and Ensemble Learning Models

Model	True Positive (TP)	True Negative (TN)	False Positive (FP)	False Negative (FN)	Accuracy (%)
Decision Tree	452	456	44	48	90.8
Random Forest	486	487	13	14	97.3
Gradient Boosting	480	484	16	20	96.4
AdaBoost	472	476	24	28	94.8
XGBoost	484	486	14	16	97

Feature Importance from the Random Forest Model

From the feature importance analysis conducted using our random forest model we identified that GFR was the most influential parameter which accounted for about 34% of model contribution. Followed closely by Creatinine Level with 27%, then BUN at 15%. In contrast to those parameters, other variables such as age and urine output account for 10% and 8% respectively while hypertension and diabetes have less significance on the model prediction where they score about 4% and 2% accordingly.

The findings reveal an essential relationship between these parameters and their impact on CKD classification based on patient data since these parameters represent significant markers associated with kidney functionality. This analysis also provides us with valuable insights into how much each factor contributes to overall decision-making process made by this machine learning model when predicting whether patients suffer from chronic kidney disease or not through providing us insight into what factors are most important when making predictions.

Table 5: Feature O\of Importance

Rank	Feature	Importance
1	GFR	0.34
2	Creatinine Level	0.27
3	BUN	0.15
4	Age	0.1
5	Urine Output	0.08
6	Hypertension	0.04
7	Diabetes	0.02

Comparison of Ensemble Learning Algorithms

As shown in Table 6, the results indicated that the random forest model had higher accuracy rate than other algorithms (97.3%), with low training time. This implies a good trade-off between the prediction performances of these models and their computational efficiency. XGBoost provides comparable prediction ability compared to other methods at 97% but requires more moderate training times. Gradient boosting achieves 96.4% accuracy rates under similar computational demands as the others, which was slower in training time when compared to random forests. Finally, we can conclude from this study that although the AdaBoost algorithm has the fastest training time with its less accurate predictions (94.8%) amongst all algorithms tested here, the random forest method gives the most efficient overall performance for ensemble machine learning models, balancing out very high levels of accuracy against significantly decreased computational requirements.

Table 6: Comparison of Ensemble Algorithms

Model	Training Time	Accuracy	Advantages
Random Forest	Low	97.30%	High accuracy, robust to overfitting
XGBoost	Medium	97.00%	Excellent predictive power
Gradient Boosting	Medium	96.40%	Good generalization
AdaBoost	Fast	94.80%	Simple and efficient

Discussion

This study compared the efficacy of five supervised machine learning systems in the prediction of chronic kidney disease (CKD) and showed that ensemble learning techniques were superior to classical Decision Tree classifiers. Out of the tested models, Random Forest proved to be the most effective one with the accuracy of 97.3%, precision of 97.1%, recall of 97.5%, F1 score of 97.3%, and ROC-AUC equal to 0.99. XGBoost has shown almost the same rate of 97.0% accuracy. The results of the present study indicate that ensemble learning methods are effective for CKD classification since they utilize multiple decision trees, thus decreasing the variance of the approach and improving generalization abilities [6,12-13]. Results of confusion matrix analysis confirmed that Random Forest is the most reliable algorithm, having demonstrated the least number of false-positive and false-negative predictions. The results of this research correlate with the findings of the previous studies claiming that Random Forest and XGBoost are more effective in medical diagnostics compared to regular decision tree models owing to their resistance to overfitting, as well as ability to detect complex nonlinear dependence among clinical variables [1-2].

The analysis of feature importance brings important clinical implications, too. Among several factors analyzed, glomerular filtration rate (GFR) appears as the strongest predictor, followed by serum creatinine level and blood urea nitrogen (BUN) [14-15]. This correlates well with nephrology literature that establishes GFR as the primary parameter of kidney performance. Also, various works acknowledge creatinine as a regular biomarker of chronic kidney disease (CKD). Langsford et al. (2017) emphasized that contribution of diabetes and hypertension is lower than one could expect [16]. This does not mean that these diseases are less important risk factors for CKD. However, their effect can become insignificant while making predictions with prediction models.

On a theoretical note, the study adds to mounting evidence that ensemble learning algorithms outperform traditional single-tree classifiers in healthcare. The comparison shows that combining learners boosts classification stability without compromising discrimination capability. Moreover, feature importance analysis contributes to making models easy to interpret, which is one of the major drawbacks in terms of the implementation of artificial intelligence solutions in healthcare. Interpretable machine

learning shows how doctors can see what factors mainly influence predictions [17-18].

In practical terms, the results imply that Random Forest can be a viable clinical decision-making tool for the early diagnosis of CKD based on commonly used laboratory and clinical indicators. The combination of high accuracy with minimal computational complexity makes it possible to integrate it into hospital IT systems and EHRs [19]. The early identification of patients at risk enables timely interventions which influence the rate of disease development, the cost of treatment connected with dialysis and kidney failure, and outcomes in patients. Furthermore, identifying GFR, creatinine, and BUN as the key predictors can help doctors focus on the most important tests in assessing the patient.

Even though the results seem quite encouraging and promising at present, it is important to point out certain limitations as well. To begin with, the present analysis relied on an only single publicly available dataset and was based on 1,000 observations, which limits the scope of application of the findings. Further, one more weakness of the research is that only seven predictor variables were considered, while other significant factors such as demographic, genetic, imaging, etc. could have increased the predictive capabilities of the study. In addition, the model was validated only on one dataset, without any external validation performed. Furthermore, although Random Forest makes it possible to get a measure of the importance of the features, even more advanced explainable AI technologies that involve SHAP and LIME were not used to provide an individual explanation for each patient [15,17].

The uniqueness of the present paper is the comparison of five supervised learning methods based on a complete and high-quality dataset of chronic kidney diseases, with no missing or duplicated records. The paper analyzes the predictive performance, confusion matrix features, computational resource consumption, and importance of input features in one framework. Considerably differing from many studies that pay attention only to the prediction, this paper considers the interpretability of the model to obtain clinically useful results. An important direction for future research is validating the obtained results on larger multicenter datasets, using deep learning methods and hybrid ensembles with wider ranges of clinical biomarkers and patient data from longitudinal studies, and employing explainable AI to increase transparency and facilitate implementation in practice.

Conclusion

Study assessed the usability of five supervised machine learning techniques in the forecast of chronic kidney disease (CKD) using seven commonly available clinical metrics. Among the reviewed algorithms, Random Forest was the best performing method recording an accuracy of 97.3%, a precision of 97.1%, a recall of 97.5%, an F1 score of 97.3%, and an ROC-AUC of 0.99. The second leading algorithm, XGBoost, was nearly equally accurate. The importance of the parameters used in the forecast was also investigated, with the results showing that glomerular filtration rate, creatinine levels, and blood urea nitrogen are the three most significant signs of the disease. This study is particularly important, as it presents an exhaustive comparison of different forms of ensemble learning algorithms applied to complete datasets without records that would not be valid nor have exact duplicates of records. The results show that Random Forest can be a useful and effective clinical decision-support technology for the early diagnosis of CKD. However, the experiment is constrained by the use of a single public dataset and a limited array of predictors. Additional research is recommended to test the models using bigger multicenter datasets, including more clinical and longitudinal biomarkers, as well as examine the possibilities of explainable artificial intelligence.

Competing Interests

The author has declared that no competing interests exist.

Data availability

Data will be made available on request.

References

1. Levin A, Ahmed SB, Carrero JJ, Foster B, Francis A, et al. (2024) Executive summary of the KDIGO 2024 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease: known knowns and known unknowns. Kidney international 105: 684-701.
2. Foundation NK (2012) KDOQI clinical practice guideline for diabetes and CKD: 2012 update. American Journal of Kidney Diseases 60: 850-886.
3. Bishop CM, Nasrabadi NM (2006) Pattern recognition and machine learning. New York: springer https://link.springer.com/book/9780387310732.
4. Breiman L (2001) Random forests. Machine learning 45: 5-32.
5. Chen T, Guestrin C (2016) Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining 785-794.
6. Chicco D, Jurman G (2020) The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC genomics 21: 6.
7. Lipton ZC (2018) The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. Queue 16: 31-57.
8. Zhang G, Song J, Qiao J (2026) Explainable AI in Deep Learning: Methods, Challenges, and Future Directions. Authorea Preprints https://www.techrxiv.org/doi/full/10.36227/techrxiv.176784518.86968822/v1.
9. Herrington WG, Judge PK, Grams ME, Wanner C (2026) Chronic kidney disease. The Lancet 407: 90-104.
10. Pérez-Pacheco Y, Parajuli D, García-Valls R (2026) Emerging Nanomedicine Strategies for Chronic Disease Management Based on Chitosan. International Journal of Molecular Sciences 27: 1387.
11. Gudsoorkar P, Lerma E, Karam S, Bencharit S, Samaranayake L, et al. (2026) Integrating Oral Health Into Kidney Care: A Policy Imperative for Chronic Kidney Disease, the US Experience. International Dental Journal 76: 109324.
12. Saad H (2020) Use bagging algorithm to improve prediction accuracy for evaluation of worker performances at a production company. arXiv preprint https://arxiv.org/pdf/2011.12343.
13. Saad H (2020) The application of data mining in the production processes. arXiv preprint https://arxiv.org/abs/2011.12348.
14. Kalantar-Zadeh K, Jafar TH, Nitsch D, Neuen BL, Perkovic V (2021) Chronic kidney disease. The lancet 398: 786-802.
15. Webster AC, Nagler EV, Morton RL, Masson P (2017) Chronic kidney disease. The lancet 389: 1238-1252.
16. Langsford D, Tang M, Cheikh Hassan HI, Djurdjev O, Sood MM, et al. (2017) The association between biomarker profiles, etiology of chronic kidney disease, and mortality. American journal of nephrology 45: 226-234.

17. Hu C, Li L, Huang W, Wu T, Xu Q, et al. (2022) Interpretable machine learning for early prediction of prognosis in sepsis: a discovery and validation study. Infectious diseases and therapy 11: 1117-1132.
18. Elshawi R, Al-Mallah MH, Sakr S (2019) On the interpretability of machine learning-based model for predicting hypertension. BMC medical informatics and decision making 19: 146.
19. Yuan S, Guo L, Xu F (2026) Artificial intelligence in nephrology: predicting CKD progression and personalizing treatment. International urology and nephrology 58: 2615-2645.