Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 6, No. 5, October 2025, Page. 3111-3123
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.5.5271

Enhancing Chronic Kidney Disease Classification Using Decision Tree And Bootstrap Aggregating: Uci Dataset Study With Improved Accuracy And Auc-Roc

Zuriati[*1], Dian Meilantika[2], Atika Arpan[3], Rizka Permata[4], Sriyanto[5], Mohd. Zaki Mas'ud[6]

[1]Department of Internet Engineering Technology, Politeknik Negeri Lampung, Indonesia
[2,3,4]Department of Information Management, Politeknik Negeri Lampung, Indonesia
[5]Department of Informatics, Institute of Informatics and Business Darmajaya, Indonesia
[6]Faculty of Artificial Intelligence and Cyber Security, Universiti Teknikal Malaysia Melaka, Malaysia

Email: [*1]zuriati@polinela.ac.id

Received : Aug 14, 2025; Revised : Sep 2, 2025; Accepted : Sep 2, 2025; Published : Oct 16, 2025

Abstract

Chronic Kidney Disease (CKD) is a progressive medical disorder that requires timely and precise identification to avoid permanent impairment of kidney function. However, Decision Tree models, although widely used in clinical applications due to their transparency, ease of implementation, and ability to handle both categorical and numerical data, are prone to overfitting and instability when applied to small or imbalanced datasets. The purpose of this study is to optimize CKD classification by integrating Bootstrap Aggregating (Bagging) with Decision Tree to enhance accuracy and robustness. The methodology involves testing two model variants a standalone Decision Tree and a Bagging-supported Decision Tree using 10-fold cross-validation and evaluating performance with accuracy, precision, recall, F1-score, and the area under the ROC curve (AUC-ROC). Findings reveal that Bagging enhances model accuracy from 0.980 to 0.987, raises precision from 0.976 to 1.000, and improves recall from 0.954 to 0.954, and increases F1-score from 0.965 to 0.976. These results demonstrate that Bagging significantly improves the reliability and generalizability of Decision Tree classifiers, making them more effective for CKD prediction.

Keywords : *Bootstrap Aggregating, Chronic Kidney Disease, Classification, Decision Tree, Ensemble Learning.*

1. INTRODUCTION

According to the KDIGO 2024 Clinical Practice Guidelines [1] Chronic Kidney Disease (CKD) is a progressive disorder characterised by a decrease in glomerular filtration rate (eGFR) of less than 60 mL/minute/1.73 m² for at least 3 months or the presence of biomarker abnormalities, such as albuminuria. CKD poses a significant global health burden, with a prevalence of 8 to 16% and caused 3.16 million deaths in 2019 [2],[3]. This burden continues to increase, especially in low- to middle-income countries, recording 434 million cases in 2021 (108 million related to type 2 diabetes) [3]. Longitudinal analysis shows an increase in incidence, mortality, and Disability-Adjusted Life Years (DALYs) since 1990, with projections that it will be the fifth leading cause of death globally by 2040 [4]. In addition, CKD also exacerbates cardiovascular disease and gout, further exacerbating its clinical and economic impact. Therefore, systematic policy interventions are urgently needed, especially in areas with low sociodemographic indices, to improve early detection and management of CKD [4]. Lai et al. [5] also emphasised this in their systematic analysis for the Global Burden of Disease (GBD) Study 2021, which showed a significant and steadily increasing global burden of CKD from 1990 to 2021. More specifically, the global burden of Chronic Kidney Disease caused by glomerulonephritis also

Vol. 6, No. 5, October 2025, Page. 3111-3123
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.5.5271

shows a significant increasing trend from 1990 to 2021, with projections to 2036, as outlined by Wang et al. [6], which also uses GBD data.

Factors like diabetes, hypertension, obesity, and the use of herbal remedies hasten the advancement of CKD, particularly in developing nations [7],[8],[9]. Timely identification is essential in averting progression to terminal kidney failure. Nevertheless, traditional diagnostic techniques, such as measuring serum creatinine concentrations and the albumin-to-creatinine ratio (ACR), often face constraints related to expense, duration, and interpretive complexity [10]. Cavalier et al. [1] highlight that early detection of CKD is possible. Nonetheless, its cause still lacks clarity, and the existence of albuminuria is still closely associated with a negative prognosis. To tackle these challenges, machine learning methods provide smart solutions due to their capability to automatically, accurately, and measurably identify clinical patterns, thereby surpassing the constraints of traditional techniques. [5],[9],[10]. Machine learning–based approaches therefore play a crucial role in providing rapid, accurate, and cost-effective diagnostic support for nephrologists, while offering interpretability that facilitates clinical trust and practical integration into electronic health records (EHR) and digital health systems, particularly in resource-limited settings.

Mahajan et al. [9] highlight the use of algorithms such as Decision Tree, Random Forests, and Support Vector Machine for various applications, including CKD stage classification, kidney transplantation, and dialysis management. Various ML algorithms have been used for the classification of CKD, both single and ensemble. Kalupukuru and Natarajan [12] evaluated multiple methods, including Decision Tree, Random Forest, and Naïve Bayes, for predicting the prognosis of CKD. Meanwhile, the effectiveness of classical algorithms (Decision Tree, K-Nearest Neighbors, Random Forest) in the early detection of CKD has been demonstrated by Ullah and Jamjoom [13] as well as Mendapara [14].

Among these algorithms, Decision Tree is a popular choice due to its ability to handle both numerical and categorical data and produce predictions that are easy to interpret [13]. Garonga et al. [15], demonstrated that Decision Tree has been effectively applied in various domains, including classification tasks with structured datasets, due to its straightforward rule-based decision paths, which can also be adapted for medical applications such as CKD classification. Beyond CKD, Decision Tree has also been successfully applied in other medical domains. For example, Biddinika et al. [16] demonstrated reasonable performance in heart disease prediction, further highlighting the adaptability of Decision Tree for healthcare classification tasks. Several studies have also shown that the Decision Tree algorithm exhibits good tolerance to noise and is capable of producing clear, rule-based decisions, making it an ideal choice for medical classification applications [17]. However, Decision Tree is prone to overfitting, especially on small, unbalanced datasets [18]. Single Decision Tree have limitations in terms of objectivity and consistency, making it challenging to replicate classification results and retest by other researchers. Additionally, Decision Tree is less capable of capturing complex relationships between features compared to more advanced models, such as artificial neural networks [19].

To overcome the weaknesses of Decision Tree, the application of ensemble techniques such as Bootstrap Aggregating (Bagging) has been proven to improve classification performance by reducing model variance, while improving its accuracy [9],[20],[21], stabilize performance [21],[22] and minimize overfitting [21],[23]. Various studies have demonstrated the effectiveness of this approach. For example, Kaur et al. [25] showed that the implementation of Bagging significantly improved the accuracy of the CKG prediction model compared to individual algorithms, such as Support Vector Machine and K-Nearest Neighbors. Meanwhile, Elshewey et al. [26] applied ensemble techniques by combining the Extra Trees Classifier with the selection of Binary Bat Swarm Search (BBSS) features to improve accuracy and interpretability, including through the explainable AI (XAI) approach in Decision Tree-based CKD classification. Recent works also reported superior performance using Random Forest

Vol. 6, No. 5, October 2025, Page. 3111-3123
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.5.5271

with AUC 0.99 [14], Extra Trees with AUC 0.998 [26], hybrid ensemble models achieving F1-scores above 0.97, and an ensemble framework integrating twelve classifiers that achieved up to 99% accuracy in CKD prediction [27]. In addition, a deep learning–based ensemble combining CNN, LSTM, and BLSTM architectures reported 98% accuracy for 6-month prediction and 97% for 12-month prediction horizons [28]. These benchmarks provide a strong reference point for evaluating the competitiveness of our proposed Bagging-Decision Tree approach.

The combination of various classification algorithms with ensemble techniques can enhance the prediction accuracy and generalisation capabilities of the model, resulting in improved performance [29]. Therefore, combining Decision Tree with Bagging is an effective strategy to maintain Decision Tree interpretability while enhancing its stability and accuracy on complex medical data. The effectiveness of the ensemble approach in predicting CKD was also strengthened by the findings of Bijoy et al. [30], which reported up to 99.5% accuracy through a combination of several algorithms, including Decision Tree, with stacking and voting. Some studies report that the integration between Decision Tree and Bagging results in superior performance over a single Decision Tree, with an accuracy of up to 97.5% [26]. Debal and Sitote [31] even found that although Random Forest and Support Vector Machine achieved the best accuracy in both binary and multiclass CKD classifications, the Decision Tree model remained competitive, with accuracies of 98.5% for binary classification and 77.5% for five-stage classification of CKD based on the severity of kidney function. The stages of CKD are generally divided into five stages based on the eGFR score, namely Stage 1 (≥90) to Stage 5 (<15), which reflect a progressive decline in kidney function [1]. Importantly, while these advanced ensembles achieve very high metrics, their "black-box" nature makes clinical deployment difficult. In contrast, Bagging with Decision Tree maintains interpretability, enabling clinicians to trace decision paths, which is crucial for integration into electronic health records (EHR) and digital health systems, particularly in resource-limited hospitals.

This study is designed to enhance the predictive performance of Chronic Kidney Disease (CKD) classification by integrating the Bagging ensemble technique with the Decision Tree algorithm. The objective is not only to improve model robustness and accuracy but also to compare its outcomes with those of a traditional single Decision Tree. The evaluation relies on multiple metrics, including accuracy, precision, recall, F1-score, and the AUC-ROC curve, ensuring a comprehensive assessment of each model's effectiveness.

2. METHOD

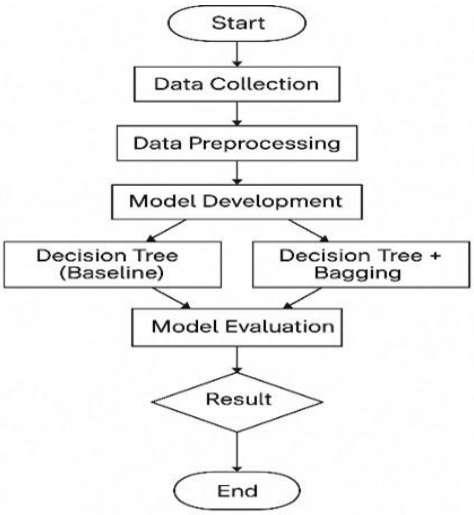

Figure 1. Research Diagram of Stages.

Vol. 6, No. 5, October 2025, Page. 3111-3123
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.5.5271

This study applies the Decision Tree algorithm combined with the Bagging method to improve Chronic Kidney Disease (CKD) classification. Two approaches are examined: the standard Decision Tree as the reference model, and a Bagging enhanced version for greater predictive robustness. A 10-fold cross-validation strategy ensures result reliability. Model performance is evaluated using common classification metrics, including accuracy, precision, recall, F1-score, and the ROC curve (AUC). In addition, figure 1 presents a research diagram of stages that illustrates the overall experimental pipeline, starting from dataset collection and preprocessing to model development and evaluation.

2.1. Dataset Collection

Table 1. Attribute Used in The CKD Dataset

No	Feature Description	Feature Name	Data Type	Measurement Unit
1	Patient's age in years	Age of the patient	Numeric	Years
2	Arterial blood pressure measurement	Blood pressure level	Continuous numerical	mmHg
3	Urine concentration relative to water density	Urine specific gravity	Ordinal (discrete scale)	1.005–1.025
4	Albumin concentration in urine	Urine albumin content	Ordered categorical	Levels 0–5
5	Glucose level detected in urine	Urine sugar presence	Ordered categorical	Levels 0–5
6	Observation of red blood cell shape	RBC morphology	Nominal (categorical)	Normal / Abnormal
7	White blood cells in urine (suggests infection)	Pus cell detection	Nominal	Normal / Abnormal
8	Pus cell clumping in urine	Clustered pus cells	Nominal	Present / Not Present
9	Presence of bacteria in urine	Bacterial infection marker	Binary indicator	Present / Not Present
10	Random glucose concentration	Random blood glucose	Numerical	mg/dL
11	Blood urea content	Urea level	Numerical	mg/dL
12	Creatinine in blood serum	Serum creatinine value	Numerical	mg/dL
13	Sodium level in blood	Sodium concentration	Numerical	mEq/L
14	Potassium level in blood	Potassium concentration	Numerical	mEq/L
15	Hemoglobin quantity in blood	Hemoglobin reading	Numerical	g/dL
16	Ratio of RBC volume in blood	Packed cell volume	Numerical	Percentage (%)
17	White blood cell count	WBC count	Numerical	Cells per mm^3
18	Red blood cell count	RBC count	Numerical	Million cells/mm^3
19	History of high blood pressure	Hypertension status	Binary (yes = 1, no = 0)	-
20	Diabetes diagnosis status	Diabetes indication	Binary (yes = 1, no = 0)	-
21	Presence of heart disease	Coronary artery disease (CAD)	Binary (yes = 1, no = 0)	-
22	Appetite status	Appetite condition	Nominal	Good / Poor
23	Swelling in legs or ankles	Pedal edema	Binary (yes = 1, no = 0)	-
24	Reduced red blood cell count	Anemia indicator	Binary (yes = 1, no = 0)	-
25	Final classification outcome	CKD status	Nominal target variable	CKD (1) / Not-CKD (0)

Vol. 6, No. 5, October 2025, Page. 3111-3123
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.5.5271

The dataset used in this study was obtained from the UCI Machine Learning Repository, titled Chronic Kidney Disease Dataset (https://archive.ics.uci.edu/dataset/336/chronic+kidney+disease). It consists 400 patient records with 24 input attributes and one target attribute (CKD/Not-CKD). The data, derived from medical and laboratory examinations, represent patients with suspected Chronic Kidney Disease. Table 1 lists the attributes, including numerical and nominal features, such as age, blood pressure, albumin levels, and haemoglobin, among others. The dataset contains missing values and format inconsistencies, which are addressed during preprocessing. It is also important to note that the dataset is imbalanced (approximately 62.5% CKD vs. 37.5% non-CKD). In this study, the Bagging approach inherently addresses this issue by resampling bootstrap subsets with replacement, thereby reducing variance and mitigating the adverse effects of imbalance while maintaining model stability.

2.2. Data Preprocessing

The preprocessing stage ensures that the dataset is clean, consistent, and in a suitable condition for use in the machine learning model training process. The first step is to handle blank values. Since their proportion is relatively small and does not significantly impact class distribution, rows containing empty values are removed from the dataset. Furthermore, categorical features such as rbc, pc, pcc, htn, and others are converted into numerical format using label encoding techniques, allowing them to be processed by classification algorithms, such as Decision Tree, that require numerical input. Although the Decision Tree algorithm does not depend on feature scaling, numerical attributes are still normalized using the Z-score method. This normalization prepares the data for exploratory statistical analysis and modeling, and helps identify characteristics of the data distribution, including potential outliers and imbalances, which will be discussed further in the Results and Discussion section. It is also important to note that the dataset is imbalanced, with a higher proportion of CKD cases. In this study, the Bagging approach helps mitigate this issue by resampling bootstrap subsets with replacement, reducing variance and minimizing the adverse effects of imbalance while maintaining model stability.

2.3. Application of Decision Tree and Bootstrap Aggregating

This research utilizes the Decision Tree algorithm as the primary classifier, with performance enhancement achieved through the application of the Bootstrap Aggregating (Bagging) method. Bagging operates by generating multiple decision trees using bootstrap samples random subsets of the training data selected with replacement. Each tree is trained independently, and the final prediction is determined by majority voting across the ensemble. Formally, the Bagging prediction for classification can be written as:

$$f_{bag}(X) = \frac{1}{B} \sum_{b=1}^{B} f_b\,(X) \qquad (1)$$

where B is the number of bootstrap samples, f_b(X) is the prediction of the b-th tree. This aggregation reduces prediction variance while maintaining interpretability.

In this study, the Bagging method was implemented using Decision Tree as the base learner, without incorporating any feature selection or hyperparameter tuning. The evaluation of model performance before and after applying Bagging was carried out using 10-fold cross-validation, and assessed through standard classification metrics: accuracy, precision, recall, F1-score, and the area under the ROC curve (AUC). For benchmarking purposes, reported performances of alternative ensemble techniques such as Random Forest [14] and XGBoost are also referenced in the discussion, though not directly implemented in this study.

Vol. 6, No. 5, October 2025, Page. 3111-3123
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.5.5271

2.4. Model Evaluation

To assess how well the models perform in classifying Chronic Kidney Disease (CKD), this study compares two primary approaches: a standard Decision Tree model and an enhanced version integrated with the Bagging ensemble technique. Each model's effectiveness was evaluated using four core classification metrics. Accuracy measures the proportion of correctly classified instances across the entire dataset. Precision reflects the model's ability to avoid false positives, while recall (or sensitivity) indicates how successfully it identifies true positive cases. The F1-score, a harmonic average of precision and recall, helps balance both aspects. To provide a deeper understanding of model performance per class, confusion matrices and detailed classification reports were also employed. These tools allow for more granular analysis beyond overall scores. In order to ensure the reliability and fairness of the evaluation, a 10-fold cross-validation scheme was applied throughout the experiments. This method helps minimize bias and variation caused by how the data is split.

Finally, the results from both models were directly compared to determine how significantly the Bagging technique improved predictive performance, particularly in terms of accuracy, sensitivity, and resistance to overfitting in CKD classification tasks.

3. RESULT

3.1. Summary of Data Preprocessing Results

The preprocessing phase is carried out to enhance the quality and reliability of the data prior to model training. During this process, question mark symbols ('?') indicating invalid values are converted into standard Not a Number (NaN) format. All rows containing missing values are then removed, as they are relatively few and do not significantly affect class distribution. Categorical features are transformed into numerical format using label encoding, while numerical features are normalized using the Z-score method to support statistical visualizations such as boxplots.

Table 2 presents sample data from the CKD dataset after all preprocessing steps have been completed. At this stage, the data has been cleaned of missing values, categorical variables have been encoded into numerical form, and numerical variables have been normalized using the Z-score method. The target class column, which indicates the presence or absence of CKD, has been separated from the feature set and is not displayed in the table. These results indicate that the dataset is ready to be used for training and evaluating classification models.

Table 2. Preprocessed Data Samples from the CKD Dataset

Age	bp	sg	al	su	rbc	pc	pcc	ba	Bgr	bu	sc	sod	pot	hemo	pcv	wc	rc	htn	dm	cad	appet	pe	ane
-0.1011	-0.3636	-2.7134	2.2735	-0.3122	1	0	1	0	-0.2215	0.0725	0.5252	-3.7301	-0.6166	-0.8657	11	42	14	1	0	0	1	1	1
0.2223	1.4317	0.0231	0.8537	-0.3122	0	0	1	0	-0.9476	1.1519	1.6335	-3.3283	-0.2703	-1.4574	8	11	12	1	1	0	1	0	1
0.8690	-0.3636	-1.8012	1.5636	-0.3122	0	0	1	0	3.8412	0.1571	0.1667	-1.0512	-0.1260	-1.0050	11	25	13	1	1	0	1	1	0
1.1923	0.5341	-1.8012	1.5636	2.1544	1	0	1	1	0.3964	0.7921	0.6230	-1.1852	0.5088	-2.8149	0	8	2	1	1	1	1	1	0

3.2. Data Distribution Visualisation

To explore the dataset characteristics, a boxplot was used to visualize the distribution of numerical features in the CKD data, as shown in figure 2. The visualization revealed that variables such as blood glucose random (bgr) and blood urea (bu) contain notable outliers, indicating high variance among patients. Serum creatinine (sc) and sodium (sod) also showed skewed distributions with inconsistent scales across observations. These irregularities reflect the clinical diversity in CKD datasets and pose

Vol. 6, No. 5, October 2025, Page. 3111-3123
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.5.5271

challenges for Decision Tree models, which are sensitive to extreme values and unstable in noise or imbalance data [16],[30]. Since traditional Decision Trees can be overly influenced by such variations, their predictive performance may become inconsistent. To address this, Bagging is introduced to enhance model stability. By combining predictions from smultiple trees built on varied bootstrap subsets, Bagging reduces variance and mitigates the influence of outliers or imbalanced distributions [24], making it effective for clinical datasets where irregularity are common.

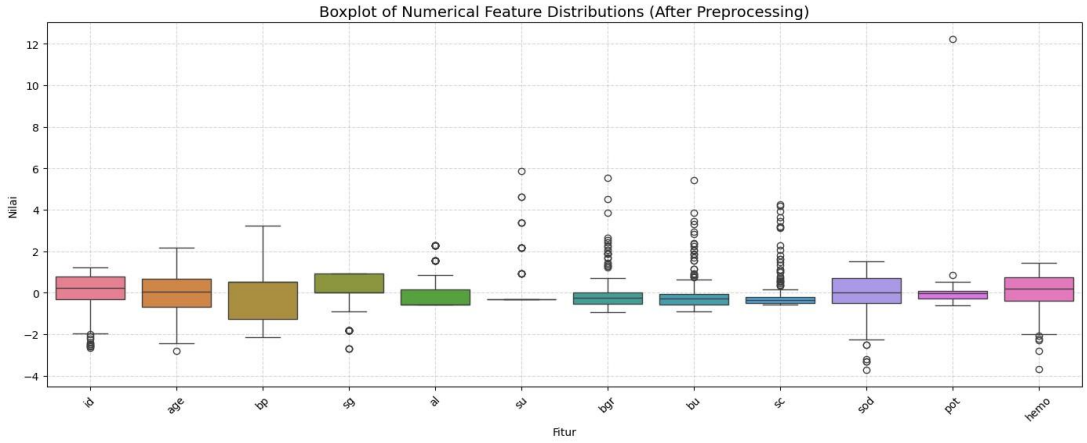

Figure 2. Boxplots of Numerical Features in the CKD Dataset

3.3. Decision Tree Baseline Model Evaluation Results

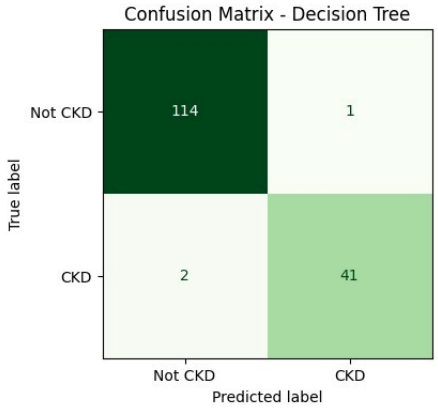

Figure 3. Confusion Matrix for Baseline Decision Tree

To evaluate the classification capability of the algorithm for Chronic Kidney Disease, a baseline model using the Decision Tree without any ensemble method was first implemented. Performance assessment through a 10-fold cross-validation process yielded an accuracy of 0.980, precision of 0.976, recall of 0.953, and F1-score of 0.964. These metrics suggest that the model is highly effective in detecting CKD cases, with a relatively low rate of classification error. As depicted in figure 3, the confusion matrix indicates that while the majority of samples were accurately classified, a small number of CKD cases were misclassified as Not-CKD.

This is reinforced by the classification report, which shows that the CKD class has a recall of 0.953, indicating that the model still fails to recognize a small number of positive cases. Meanwhile, the Not-CKD class achieves a higher recall of 0.991, demonstrating more consistent performance in identifying negative cases. The precision values for both classes are strong, with 0.976 for CKD and

Vol. 6, No. 5, October 2025, Page. 3111-3123
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.5.5271

0.983 for Not-CKD, resulting in an F1-score of 0.964 for the CKD class and 0.987 for the Not-CKD class. The model also attains an overall accuracy of 0.981, with macro and weighted average scores for all metrics consistently ranging from 0.972 to 0.979.

These results suggest that although the default Decision Tree model performs well overall, it still shows a mild preference for the majority class, which reflects a slight class imbalance in the dataset. Therefore, ensemble techniques such as Bagging are recommended to further enhance the model's generalization performance and reduce potential bias toward the minority class.

3.4. Bagging and Decision Tree Model Evaluation Results

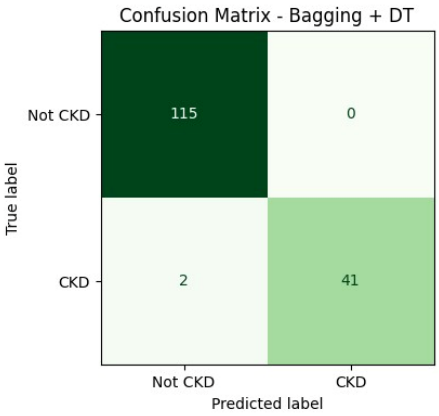

Figure 4. Confusion Matrix for Bagging and Decision Tree

The application of the Bagging ensemble method to the Decision Tree algorithm led to a notable enhancement in classifying Chronic Kidney Disease (CKD) cases. Evaluation using the 10-fold cross-validation approach showed that the Bagging model achieved 0.987 accuracy, indicating high predictive performance. For CKD cases (class 1), the model attained 1.000 precision, 0.953 recall, and an F1-score of 0.976. Meanwhile, for Not-CKD cases (class 0), it achieved 0.983 precision, 1.000 recall, and an F1-score of 0.991. These results reflect the model's excellent capability to correctly identify both classes, with no false negatives in the Not-CKD class and only a minimal number of misclassifications in the CKD class. These outcomes are visually represented in the confusion matrix shown in figure 4.

3.5. Comparative Analysis of Model Performance

A comparative analysis was conducted to assess the impact of applying the Bagging technique on the performance of the Decision Tree model in classifying Chronic Kidney Disease. As visually represented in figure 5, the Bagging model demonstrates superior performance across key metrics. The Bagging model achieves an accuracy of 0.987, surpassing the standard Decision Tree model's accuracy of 0.980. This improvement demonstrates that Bagging makes a positive contribution to the overall accuracy of predictions.

The precision performance for the CKD class has also increased significantly with the use of Bagging. While the standard Decision Tree model yielded a precision of 0.976 for the CKD class, the Bagging model achieved a perfect 1.000 precision. This crucial improvement means that all optimistic predictions generated by the Bagging model are indeed CKD cases, effectively eliminating false positives (reducing them from 1 in the standard Decision Tree to 0 in Bagging). Meanwhile, the recall for the CKD class remained consistent at 0.953 for both models, indicating that the number of undetected

Vol. 6, No. 5, October 2025, Page. 3111-3123
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.5.5271

CKD cases (false negatives) remained the same. The increase in F1-score from 0.964 to 0.976 further illustrates that the Bagging model achieves a better balance between precision and recall.

Overall, these results, visually reinforced by figure 5, demonstrate that the Bagging technique enhances prediction accuracy, particularly in reducing Type I errors (false positives), while also providing model stability under data variability. Figure 5 clearly shows that the Bagging model consistently scores higher on all key metrics, particularly precision and F1-score. This visualization confirms the advantages of the ensemble approach in improving classification performance, especially in terms of predictive balance and resilience to uneven data distribution.

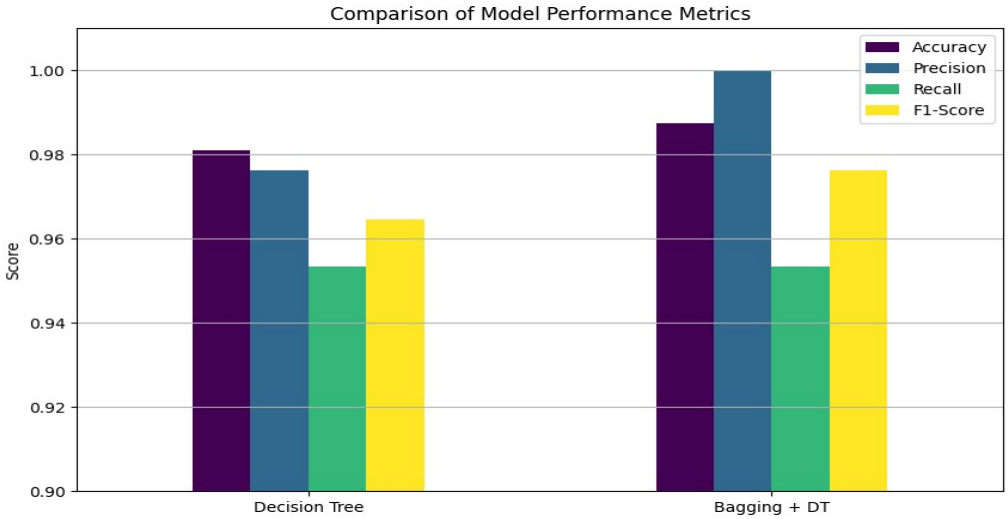

Figure 5. Comparative Bar Plot of Model Performance Metrics

Figure 6 illustrates the Receiver Operating Characteristic (ROC) curves for both the standalone Decision Tree model and the Bagging enhanced Decision Tree model. This curve represents the trade-off between the True Positive Rate (Sensitivity) and the False Positive Rate across a range of classification thresholds.

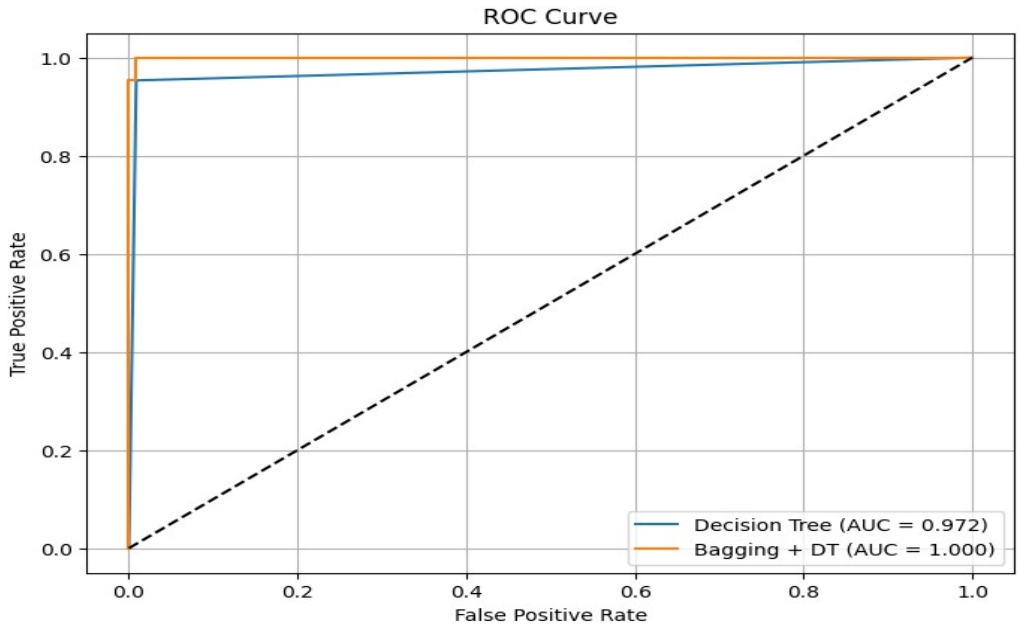

Figure 6. ROC Curve

Vol. 6, No. 5, October 2025, Page. 3111-3123
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.5.5271

Based on the visualisation, the standalone Decision Tree model achieved an Area Under the Curve (AUC) score of 0.972, indicating strong classification capability. In comparison, the Bagging + Decision Tree model reached an AUC of 1.000, reflecting improved discriminative power through the ensemble technique. A curve approaching the top-left corner signifies better separation between classes. These findings suggest that integrating Bagging into the Decision Tree approach enhances generalisation and predictive performance, particularly in addressing overfitting and high variance.

4. DISCUSSIONS

The experimental results highlight the advantages of integrating the Bagging ensemble technique with the Decision Tree algorithm in the classification of Chronic Kidney Disease (CKD). The superior performance, evidenced by an accuracy of 0.987, 1.000 precision for the CKD class, and an AUC score of 1.000, demonstrates that Bagging enhances model robustness, stability, and predictive reliability. These findings align with Debal and Sitote [31] and Moreno-Sanchez [22], who reported similar improvements through ensemble-based approaches.

By reducing variance and mitigating overfitting, a common issue in single-tree models, Bagging enables better generalization, especially for noisy or imbalanced datasets. This is crucial in early CKD detection, where minimizing false negatives and false positives has direct clinical consequences. Compared to the baseline Decision Tree, the Bagging-DT model achieved lower false positive rates and consistent sensitivity, making it more suitable for deployment in decision support systems or automated screening tools in primary healthcare.

Despite these encouraging results, this study has several limitations. The dataset, though rich in features, is relatively small and may not reflect the full clinical diversity of real-world populations. Moreover, the model was validated only on internal data; external validation with independent datasets is necessary to assess generalizability. The dataset from Kaggle, while well-structured, may not fully capture the variability, demographic diversity, and diagnostic nuances of clinical settings, which should be considered when interpreting applicability.

Future research should use larger, multi-institutional datasets and real-time clinical data such as electronic health records (EHRs). Exploring advanced ensemble strategies boosting, stacking, or hybrid frameworks along with improved feature engineering and hyperparameter tuning may further refine performance and interpretability. Beyond technical improvements, evaluating integration into clinical workflows is essential, including healthcare professional acceptance, impact on decision-making, and the potential to slow CKD progression in at-risk populations.

Practically, the Bagging-enhanced Decision Tree model offers advantages in low-resource healthcare environments. Its interpretability, computational efficiency, and high sensitivity make it suitable for rapid CKD screening in primary care, particularly where advanced laboratory facilities are limited. As an early-warning mechanism, it could prompt timely referrals for confirmatory testing and intervention, potentially improving outcomes and reducing the long-term economic burden on healthcare systems.

5. CONCLUSION

This study evaluated Chronic Kidney Disease (CKD) classification using the Decision Tree algorithm and the Bagging ensemble method. The standard Decision Tree achieved 0.980 accuracy, 0.976 precision, 0.953 recall, and 0.964 F1-score, with several misclassifications in the confusion matrix. With Bagging, performance improved to 0.987 accuracy, 1.000 precision for CKD, 0.953 recall, and 0.976 F1-score. The confusion matrix showed complete elimination of false positives and a marked reduction in false negatives crucial in clinical diagnosis, it is important to acknowledge the potential risk of overfitting associated with a perfect precision score. The Bagging-DT model also achieved a perfect

Vol. 6, No. 5, October 2025, Page. 3111-3123
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.5.5271

AUC score of 1.000, demonstrating strong discriminative power between CKD and Not-CKD. These results confirm that combining Decision Trees with Bagging enhances predictive performance, robustness, and generalization, especially with complex, imbalanced medical datasets. This study underscores the value of interpretable models in decision-support systems and how ensembles mitigate overfitting and data variability, reinforcing the importance of explainable AI (XAI) in healthcare. Future work should explore other ensemble techniques, integrate advanced feature selection and hyperparameter optimization, and validate the model with large-scale, external datasets for scalability and generalizability.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest, either among the authors or with respect to the research object presented in this paper.

REFERENCES

[1] E. Cavalier, T. Zima, P. Datta, and K. Makris, "Recommendations for European laboratories based on the KDIGO 2024 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease," *Clinical Chemistry and Laboratory Medicine (CCLM)*, vol. 63, no. 3, pp. 525–534, Feb. 2025, doi: 10.1515/cclm-2024-1082.

[2] C. Ke, J. Liang, M. Liu, S. Liu, and C. Wang, "Burden of chronic kidney disease and its risk-attributable burden in 137 low-and middle-income countries, 1990–2019: results from the global burden of disease study 2019," *BMC Nephrology*, vol. 23, no. 1, pp. 1–12, 2022, doi: 10.1186/s12882-021-02597-3.

[3] J. Guo, Z. Liu, P. Wang, and H. Wu, "Global, regional, and national burden inequality of chronic kidney disease, 1990–2021: a systematic analysis for the global burden of disease study 2021," *Frontiers in Medicine*, vol. 11, no. January, pp. 1–25, Jan. 2025, doi: 10.3389/fmed.2024.1501175.

[4] B. Bikbov, C. A. Purcell, A. S. Levey, and M. Smith, "Global, regional, and national burden of chronic kidney disease, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017," *The Lancet*, vol. 395, no. 10225, pp. 709–733, Feb. 2020, doi: 10.1016/S0140-6736(20)30045-3.

[5] Y. Lai, H. Long, Z. Liang, and C. Wu, "Global burden and risk factors of chronic kidney disease due to hypertension in adults aged 20 plus years, 1990–2021," *Frontiers in Public Health*, vol. 13, no. May, pp. 1–12, May 2025, doi: 10.3389/fpubh.2025.1503837.

[6] X. Wang, Z. Liu, N. Yi, and L. Li, "The global burden of chronic kidney disease due to glomerulonephritis: trends and predictions," *International Urology and Nephrology*, Mar. 2025, doi: 10.1007/s11255-025-04440-2.

[7] A. Makmun, B. Satirapoj, D. G. Tuyen, and M. W. Y. Foo, "The burden of chronic kidney disease in Asia region: a review of the evidence, current challenges, and future directions.," *Kidney research and clinical practice*, vol. 44, no. 3, pp. 411–433, 2024, doi: 10.23876/j.krcp.23.194.

[8] N. M. Hustrini, E. Susalit, F. F. Widjaja, and A. I. Khumaedi, "The Etiology of Advanced Chronic Kidney Disease in Southeast Asia: A Meta-analysis," *Journal of Epidemiology and Global Health*, vol. 14, no. 3, pp. 740–764, 2024, doi: 10.1007/s44197-024-00209-5.

[9] P. Mahajan, S. Uddin, F. Hajati, and M. A. Moni, "Ensemble Learning for Disease Prediction: A Review," *Healthcare*, vol. 11, no. 12, p. 1808, Jun. 2023, doi: 10.3390/healthcare11121808.

[10] F. Keshvari-Shad, S. Hajebrahimi, and M. P. Laguna Pes, "A Systematic Review of Screening Tests for Chronic Kidney Disease: An Accuracy Analysis," *Galen Medical Journal*, vol. 9, p. e1573, 2020, doi: 10.31661/gmj.v9i0.1573.

[11] B. Zhou, J. Zhang, and G. Li, "The Global, Regional, and National Burden of Pancreatitis From 1990 to 2021: A Systematic Analysis for the Global Burden of Disease Study 2021," *Journal of Gastroenterology and Hepatology*, vol. 40, no. 5, pp. 1297–1306, May 2025, doi: 10.1111/jgh.16906.

[12] S. R. Kalupukuru and K. Natarajan, "Machine Learning Methods for Predicting the Prognosis of

Vol. 6, No. 5, October 2025, Page. 3111-3123
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.5.5271

Chronic Kidney Disease," *Procedia Computer Science*, vol. 258, pp. 1372–1382, 2025, doi: 10.1016/j.procs.2025.04.370.

[13] Z. Ullah and M. Jamjoom, "Early Detection and Diagnosis of Chronic Kidney Disease Based on," *Journal of Healthcare Engineering*, vol. 2023, 2023.

[14] K. Mendapara, "Development and evaluation of a chronic kidney disease risk prediction model using random forest," *Frontiers in Genetics*, vol. 15, no. June, pp. 1–11, Jun. 2024, doi: 10.3389/fgene.2024.1409755.

[15] M. Garonga and Rita Tanduk, "Comparison of Naive Bayes, Decision Tree, and Random Forest Algorithms in Classifying Learning Styles of Universitas Kristen Indonesia Toraja Students," *Jurnal Teknik Informatika (Jutif)*, vol. 4, no. 6, pp. 1507–1514, 2023, doi: 10.52436/1.jutif.2023.4.6.1020.

[16] M. K. Biddinika, A. Masitha, H. Herman, and V. A. N. Fatimah, "Machine Learning Techniques for Heart Disease Prediction Using a Multi-Algorithm Approach," *JUITA: Jurnal Informatika*, vol. 12, no. 2, p. 149, 2024, doi: 10.30595/juita.v12i2.24153.

[17] G. Blanc, J. Lange, A. Malik, and L. Y. Tan, "Popular decision tree algorithms are provably noise tolerant," *Proceedings of Machine Learning Research*, vol. 162, pp. 2091–2106, 2022, doi: https://doi.org/10.48550/arXiv.2206.08899.

[18] M. Alizade-harakiyan, A. Khodaei, A. Yousefi, H. Zamani, and A. Mesbahi, "Decision tree-based machine learning algorithm for prediction of acute radiation esophagitis," vol. 42, no. January, 2025, [Online]. Available: https://doi.org/10.1016/j.bbrep.2025.101991

[19] D. Bertsimas and V. Digalakis, "Improving Stability in Decision Tree Models," pp. 1–37, 2023, [Online]. Available: http://arxiv.org/abs/2305.17299

[20] M. M. M. Hassan, M. M. M. Hassan, S. Mollick, and M. A. R. Khan, "A Comparative Study, Prediction and Development of Chronic Kidney Disease Using Machine Learning on Patients Clinical Records," *Human-Centric Intelligent Systems*, vol. 3, no. 2, pp. 92–104, 2023, doi: 10.1007/s44230-023-00017-3.

[21] S. Y. Irianto, D. Linda, I. I. M. Rizki, S. Karnila, and D. Yuliawati, "Chronic Kidney Disease Classification Using Bagging and Particle Swarm Optimization Techniques," *International Journal of Advanced Computer Science and Applications*, vol. 16, no. 3, pp. 689–698, 2025, doi: 10.14569/IJACSA.2025.0160368.

[22] P. A. Moreno-Sánchez, "Chronic Kidney Disease Early Diagnosis Enhancing by using Data Mining Classification and Features Selection," pp. 460–467, 2021, doi: https://doi.org/10.1007/978-3-030-69963-5_5.

[23] P. A. Moreno-Sanchez, "Data-Driven Early Diagnosis of Chronic Kidney Disease: Development and Evaluation of an Explainable AI Model," *IEEE Access*, vol. 11, no. Ml, pp. 38359–38369, 2023, doi: 10.1109/ACCESS.2023.3264270.

[24] S. Pal, "Chronic Kidney Disease Prediction Using Machine Learning Techniques," *Biomedical Materials and Devices*, vol. 1, no. 1, pp. 534–540, 2023, doi: 10.1007/s44174-022-00027-y.

[25] C. Kaur, M. S. Kumar, A. Anjum, M. B. Binda, M. R. Mallu, and M. S. Al Ansari, "Chronic Kidney Disease Prediction Using Machine Learning," *Journal of Advances in Information Technology*, vol. 14, no. 2, pp. 384–391, 2023, doi: 10.12720/jait.14.2.384-391.

[26] A. M. Elshewey, E. Selem, and A. H. Abed, "Improved CKD classification based on explainable artificial intelligence with extra trees and BBFS," *Scientific Reports*, vol. 15, no. 1, p. 17861, May 2025, doi: 10.1038/s41598-025-02355-7.

[27] D. Chhabra, M. Juneja, and G. Chutani, "An Efficient Ensemble-based Machine Learning approach for Predicting Chronic Kidney Disease," *Current Medical Imaging Reviews*, vol. 20, pp. 1–12, 2023, doi: 10.2174/1573405620666230508104538.

[28] D. Saif, A. M. Sarhan, and N. M. Elshennawy, "Early prediction of chronic kidney disease based on ensemble of deep learning models and optimizers," *Journal of Electrical Systems and Information Technology*, vol. 11, no. 1, 2024, doi: 10.1186/s43067-024-00142-4.

[29] S. Imran Ali, B. Ali, J. Hussain, M. Hussain, and F. A. Satti, "Cost-Sensitive Ensemble Feature Ranking and Automatic Threshold Selection for Chronic Kidney Disease Diagnosis," *Applied Sciences*, vol. 10, no. 16, p. 5663, Aug. 2020, doi: 10.3390/app10165663.

[30] M. H. I. Bijoy, M. J. Mia, M. M. Rahman, M. S. Arefin, P. K. Dhar, and T. Shimamura, *A robot*

Vol. 6, No. 5, October 2025, Page. 3111-3123
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.5.5271

process automation based mobile application for early prediction of chronic kidney disease using machine learning, vol. 7, no. 6. Springer International Publishing, 2025. doi: 10.1007/s42452-025-06980-9.

[31] D. A. Debal and T. M. Sitote, "Chronic kidney disease prediction using machine learning techniques," *Journal of Big Data*, vol. 9, no. 1, 2022, doi: 10.1186/s40537-022-00657-5.

[32] S. C. Tan, "Decision Tree Regression with Residual Outlier Detection," *Journal of Data Science and Intelligent Systems*, vol. 3, no. 2, pp. 126–135, 2024, doi: 10.47852/bonviewjdsis42023861.