ISSN (e): 2250 – 3005 || Volume, 16 || Issue, 9|| September – 2026 ||
International Journal of Computational Engineering Research (IJCER)

# Comparative Evaluation of Random Forest and Decision Tree Classifiers for Chronic Kidney Disease Prognosis

[1*]Aboluwoye Olumuyia Chris, [2]Arome Gabriel Junior, [3]Balogun Temitayo Elijah, [4]Babafemi Tuyi Kolade

[1]*Department of Computer Science, Federal University of Technology, Akure*
[2]*Department of Cybersecurity, Federal University of Technology, Akure*
[3]*Department of Information Systems, Federal University of Technology, Akure*
[4]*Center for Renewable Energy, Federal University of Technology, Akure*

**ABSTRACT**

*Chronic kidney disease (CKD) is a progressive condition that often develops silently before reaching an advanced stage, making early and reliable prognosis essential for reducing its clinical and economic burden. This study presents a comparative evaluation of two widely used supervised learning algorithms, Decision Tree and Random Forest, for predicting CKD status from routine clinical and demographic attributes. A publicly available Kaggle CKD dataset comprising 1,659 patient records was pre-processed using K-Nearest Neighbour imputation to handle missing values and the Synthetic Minority Oversampling Technique (SMOTE) to correct class imbalance prior to model training. Both classifiers were trained and evaluated on an identical train-test split using accuracy, precision, recall, F1-score, and confusion-matrix analysis. The Random Forest classifier achieved a notably higher classification accuracy of 91.6%, compared with 82.2% for the single Decision Tree. The findings support the adoption of ensemble tree-based methods over single-tree classifiers as a computationally inexpensive yet interpretable option for clinical decision-support systems targeting CKD screening in resource-constrained healthcare settings*

***KEYWORDS:*** *Chronic Kidney Disease; Decision Tree; Random Forest; Machine Learning; Classification; SMOTE; Clinical Decision Support*

---

Date of Submission: 25-08-2026 Date of acceptance: 05-09-2026

---

## I. INTRODUCTION

Chronic kidney disease (CKD) has become one of the fastest-growing non-communicable diseases worldwide. Recent global estimates place the age-standardised prevalence of CKD among adults aged 20 years and above at approximately 14%, corresponding to nearly 800 million people, with the disease now ranked as the ninth leading cause of death globally and one of the few major causes of mortality whose burden continues to rise rather than fall [1]. The condition is closely linked to diabetes mellitus, hypertension, and obesity, and its early stages are frequently asymptomatic, so that a large proportion of affected individuals are diagnosed only after substantial and often irreversible loss of kidney function has occurred [1].

Conventional CKD diagnosis relies on the clinical interpretation of biochemical markers such as serum creatinine, blood urea, and estimated glomerular filtration rate. While these markers remain clinically indispensable, manual interpretation does not easily scale to the volume of data now routinely captured in electronic health records, nor does it readily capture non-linear interactions among the many demographic, laboratory, and lifestyle variables that jointly influence disease progression. This gap has motivated a growing body of work applying machine learning to CKD screening and prognosis, since learning algorithms can be trained to recognise complex, non-obvious patterns in structured clinical data and to flag high-risk patients for closer clinical follow-up [8, 9].

Among the algorithms explored in the literature, tree-based methods are particularly attractive for clinical applications because they operate directly on mixed numerical and categorical data with little pre-processing, and their decision logic can, in principle, be inspected and explained to a clinician. The Decision Tree is the simplest member of this family: a single tree that recursively partitions the feature space using rules that maximise information gain at each split [5]. Although easy to interpret, a single tree is highly sensitive to the particular sample it is trained on and tends to overfit, producing rules that generalise poorly to new patients. Random Forest was introduced to address this weakness by growing many decision trees on bootstrap-

resampled subsets of the training data and random subsets of the available features, then combining their individual predictions by majority vote [4]. The resulting ensemble is generally far less prone to overfitting than any one of its constituent trees, at a modest cost to direct interpretability.

This study isolates these two related but architecturally distinct classifiers, a single Decision Tree and a Random Forest ensemble built from many such trees, and compares their ability to prognose CKD status from a clinical dataset sourced from Kaggle [2]. The comparison is designed to quantify, on a common dataset and evaluation protocol, the practical benefit that ensembling confers over a single interpretable tree, and to discuss the implications of that benefit for deploying tree-based screening tools in real clinical workflows.

## II. RELATED WORK

A substantial body of research has applied machine learning to CKD prediction using the UCI CKD dataset [3] and, more recently, synthetic or aggregated datasets hosted on platforms such as Kaggle. Islam and Logofătu [8] compared Decision Tree, Random Forest, and Naive Bayes on the UCI CKD dataset after KNN imputation and SMOTE balancing, and reported that Random Forest achieved the highest accuracy at 96%, ahead of Decision Tree at 94% and Naive Bayes at 90%, a margin broadly consistent in direction with, though smaller in magnitude than, the gap observed in the present study. Similarly, Nguyen et al. [9] evaluated six algorithms, including Random Forest, XGBoost, and Naive Bayes, on the UCI dataset and found Random Forest among the top performers alongside XGBoost, reinforcing the general pattern in this literature that ensemble tree methods tend to sit at or near the top of the accuracy ranking, while single trees and linear models trail behind.

Across this literature, ensemble methods such as Random Forest, Gradient Boosting, and XGBoost are consistently reported to outperform single classifiers such as Decision Tree and Logistic Regression on accuracy, recall for the disease-positive class, and robustness to noisy or imbalanced clinical data. This advantage is usually attributed to the variance-reduction effect of bagging and to the ability of ensembles to average out the idiosyncratic errors that any one tree makes when trained on a finite clinical sample [4].

At the same time, several studies caution that reported performance is highly sensitive to how missing data and class imbalance are handled prior to training, since medical datasets frequently under-represent the disease-positive class and contain incomplete laboratory records [8, 9]. Imputation strategies such as K-Nearest Neighbour imputation and resampling strategies such as SMOTE [6] are widely adopted to address these issues before model fitting, and the present study follows this established practice so that the comparison between Decision Tree and Random Forest reflects genuine algorithmic differences rather than artefacts of unequal pre-processing.

## III. MATERIALS AND METHODS

### *3.1 Dataset*

This study uses a publicly available synthetic CKD dataset obtained from Kaggle [2], comprising detailed health records for 1,659 patients. The dataset combines demographic details, lifestyle factors, medical history, clinical measurements (including blood pressure, blood glucose, haemoglobin, serum creatinine, blood urea, and electrolyte concentrations), medication usage, symptoms, and quality-of-life indicators, alongside a binary target label indicating CKD status. As is typical of real-world clinical data, the raw dataset contained missing values and an imbalance between CKD-positive and CKD-negative cases, as shown in Figure 1.

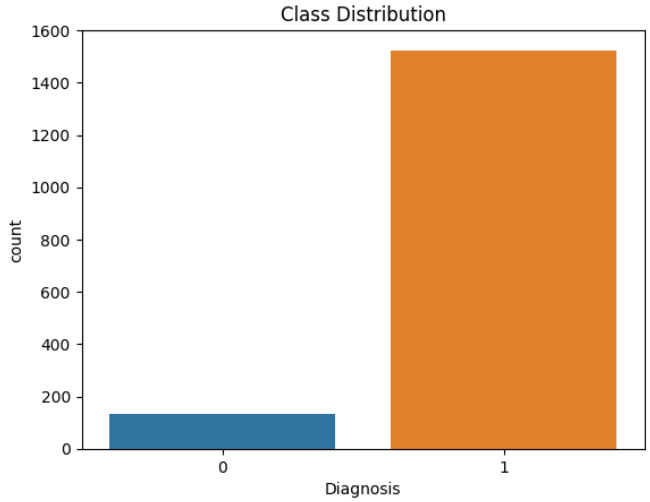


**Figure 1: Class distribution of the Kaggle CKD dataset prior to balancing.**

Figure 2 shows the pairwise correlation structure among the numerical clinical features. Most feature pairs show weak linear correlation, with the notable exception of glomerular filtration rate (GFR), which shows the strongest linear association with the CKD diagnosis label, consistent with its established clinical role as a direct indicator of kidney function.

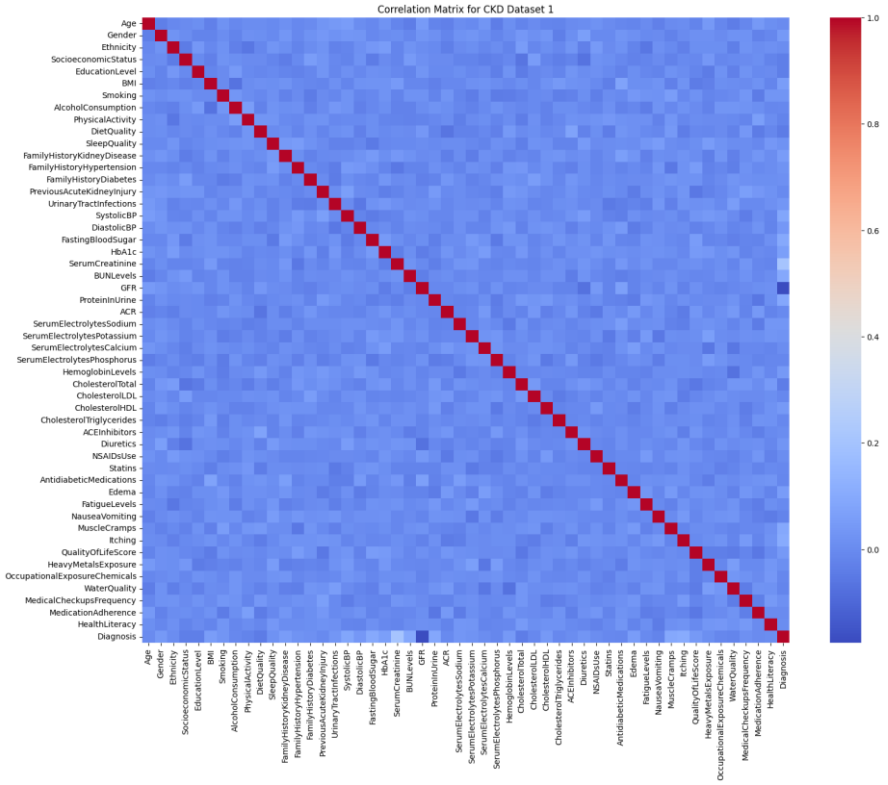


**Figure 2: Correlation matrix of clinical features in the Kaggle CKD dataset.**

### *3.2 Data Pre-processing*

Missing values were imputed using the K-Nearest Neighbour (KNN) imputation method, which estimates a missing entry from the average of the k most similar complete records rather than a simple global mean, preserving local structure in the data. Categorical variables were numerically encoded, and continuous features were scaled where required by the downstream algorithm. To correct the imbalance between the two outcome classes, SMOTE [6] was applied to the training partition only, generating synthetic minority-class samples by interpolating between existing minority instances and their nearest neighbours. In this dataset, the training partition contained 108 CKD-negative and 1,219 CKD-positive records before balancing; after SMOTE, both classes were represented equally at 1,219 samples each, removing the majority-class bias that would otherwise dominate training.

### *3.3 Classification Algorithms*

Decision Tree: A Decision Tree classifier partitions the feature space recursively, at each node choosing the split that yields the greatest reduction in class impurity, commonly measured through entropy or information gain [5]. Its principal strength is transparency, since the resulting rule set can be traced and interpreted node by node; its principal weakness is a tendency to grow overly complex trees that fit noise in the training sample, resulting in high variance and weaker performance on unseen patients.

Random Forest: A Random Forest is an ensemble of many Decision Trees, each trained on an independent bootstrap sample of the training data and permitted to consider only a random subset of features at each split [4]. The final prediction for a given patient is obtained by aggregating the votes of all trees in the forest. This combination of bootstrap aggregation (bagging) and feature randomisation decorrelates the individual trees, so that their errors tend to cancel out on average rather than compound, giving the forest lower variance and typically better generalisation than any single constituent tree.

### *3.4 Evaluation Metrics and Experimental Protocol*

Both classifiers were trained and evaluated under identical conditions using the same SMOTE-balanced training partition and held-out test partition of the Kaggle dataset, so that any performance difference can be attributed to

the algorithms themselves rather than to differences in data handling. Performance was assessed using accuracy, precision, recall, F1-score, and the confusion matrix [7], providing both an overall measure of correctness and a more detailed view of each model's ability to correctly identify CKD-positive patients, which is the clinically consequential class.

## IV. RESULTS AND DISCUSSION

Table 1 summarises the overall classification accuracy obtained by the Decision Tree and Random Forest classifiers on the held-out test partition (n = 332) of the Kaggle CKD dataset.

*Table 1: Accuracy of Decision Tree and Random Forest on the dataset.*

| **Model** | **Accuracy** |
|---|---|
| Decision Tree | 0.8223 (82.2%) |
| Random Forest | 0.9157 (91.6%) |

Table 2 breaks this down by class. Because the raw test set retains the natural class imbalance (27 CKD-negative and 305 CKD-positive cases), performance on the minority (non-CKD) class is the more demanding and clinically informative comparison.

*Table 2: Per-class precision, recall, and F1-score on the test set.*

| **Model** | **Class** | **Precision** | **Recall** | **F1-score** | **Support** |
|---|---|---|---|---|---|
| Decision Tree | Non-CKD (0) | 0.15 | 0.26 | 0.19 | 27 |
| | CKD (1) | 0.93 | 0.87 | 0.90 | 305 |
| Random Forest | Non-CKD (0) | 0.40 | 0.07 | 0.12 | 27 |
| | CKD (1) | 0.92 | 0.99 | 0.96 | 305 |

Figures 3 and 4 present the corresponding confusion matrices, and Figure 5 summarises the four headline metrics side by side.

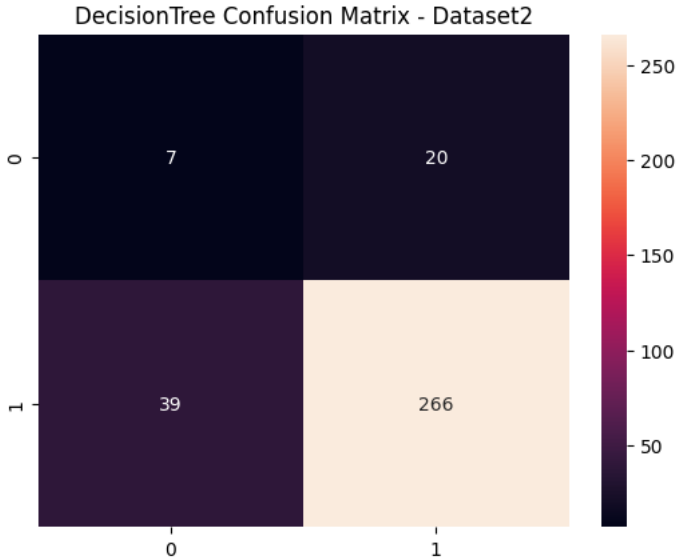


**Figure 3: Confusion matrix of the Decision Tree classifier on the Kaggle test set.**

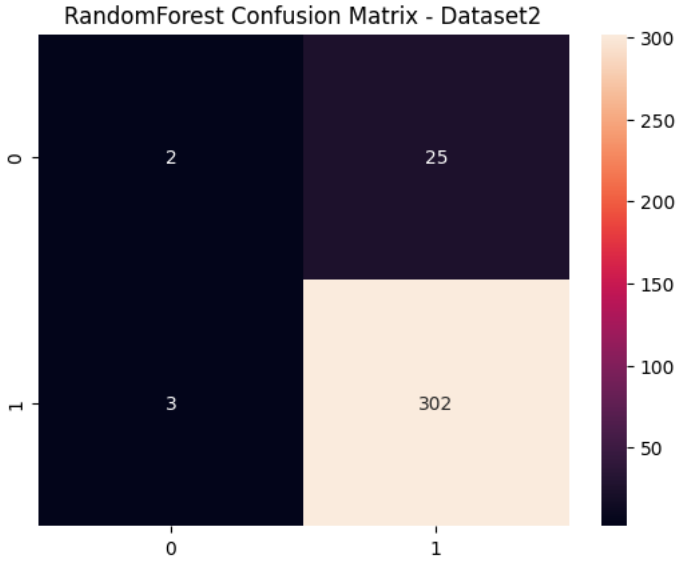


**Figure 4: Confusion matrix of the Random Forest classifier on the Kaggle test set.**

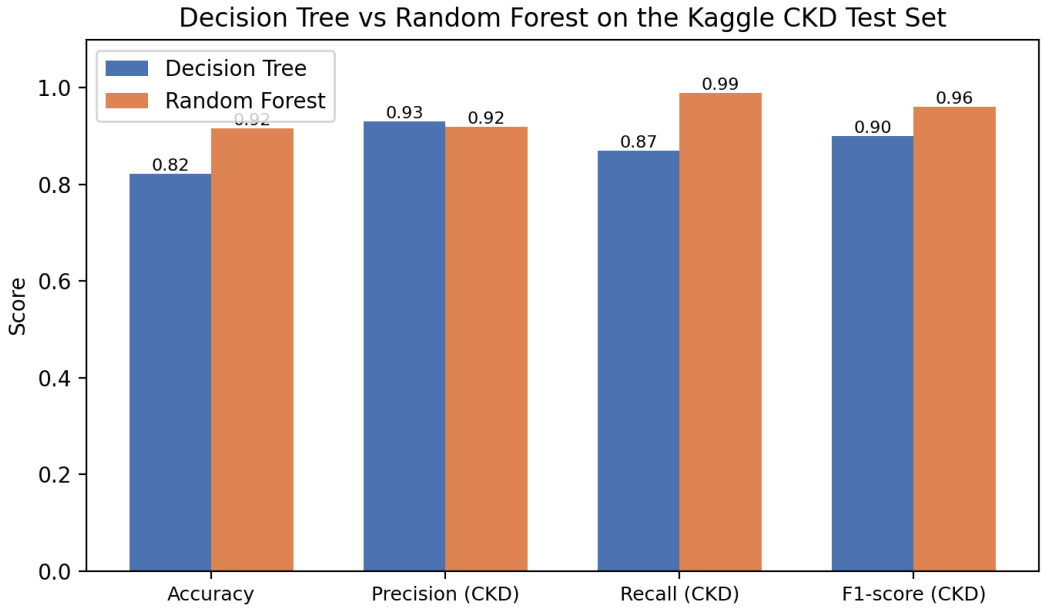


**Figure 5: Accuracy, precision, recall, and F1-score of Decision Tree versus Random Forest on the CKD-positive class**.

The Random Forest classifier outperformed the single Decision Tree by close to nine percentage points in overall accuracy (Table 1) and, more importantly for a screening application, achieved substantially higher recall on the CKD-positive class (0.99 versus 0.87, Table 2), meaning it missed far fewer true CKD cases. This gap is consistent with the theoretical expectation that ensembling reduces variance [4]: because each tree in the forest is trained on a different bootstrap sample and a randomly restricted feature set, the trees make partially independent errors, and combining their votes cancels out much of the noise that any single tree would otherwise carry into its predictions. The Decision Tree, by contrast, was trained once on the full feature set and training sample, leaving it more exposed to whatever idiosyncrasies existed in that particular training partition, which is the classic overfitting failure mode of unpruned single trees [5].

Both models struggled on the minority non-CKD class, as the confusion matrices in Figures 3 and 4 show: the Decision Tree correctly flagged 7 of 27 non-CKD patients while Random Forest correctly flagged only 2 of 27. This is a direct consequence of the residual imbalance in the untouched test set and mirrors a limitation reported elsewhere in the CKD-prediction literature, where SMOTE improves training-time balance but test-time performance on a naturally rare class remains comparatively weak [8, 9]. This weakness on the

non-CKD class is precisely where the Decision Tree's marginally higher non-CKD recall (0.26 versus 0.07) becomes clinically relevant despite its lower overall accuracy, underlining that accuracy alone can obscure important differences in how two models fail.

From a clinical-deployment perspective, the two models occupy different points on the interpretability-accuracy trade-off. The Decision Tree offers a fully transparent, human-traceable rule path from patient attributes to predicted outcome, which can be valuable where clinicians require an explicit justification for a screening flag. The Random Forest sacrifices some of that direct traceability, since its prediction is an aggregate over many trees rather than a single rule path, but this can be partially mitigated using feature-importance scores derived from the forest, which indicate which clinical attributes contribute most to its predictions without requiring the full ensemble to be inspected tree by tree. Given that the accuracy and CKD-recall gains from Random Forest are substantial while feature-importance analysis preserves a useful degree of explainability, the results favour Random Forest as the more suitable candidate for a CKD screening support tool, with the Decision Tree retained mainly as an interpretable baseline or as an illustrative model for clinician training.

It is worth noting that both models were trained and evaluated on the same SMOTE-balanced Kaggle dataset, so the reported gap reflects a genuine algorithmic difference rather than an artefact of unequal class distributions. Nonetheless, because the Kaggle dataset used here is synthetic, external validation on real-world CKD registries would be a necessary next step before either model is considered for clinical deployment, in line with recommendations made in comparable CKD prediction studies [8, 9].

## V. CONCLUSION

This study compared a single Decision Tree classifier against a Random Forest ensemble for the prognosis of Chronic Kidney Disease using a Kaggle-sourced clinical dataset, following K-Nearest Neighbour imputation and SMOTE-based class balancing. The Random Forest classifier achieved a substantially higher accuracy (91.6%) and CKD-recall (0.99) than the single Decision Tree (82.2% accuracy, 0.87 recall), confirming that bagging and feature-randomised ensembling meaningfully improve generalisation over a single tree for this prediction task. These findings, consistent with related comparative studies on other CKD cohorts, reinforce the broader case for favouring ensemble tree-based methods over single-tree classifiers in clinical decision-support applications, while recognising that Decision Trees retain value where full rule-level transparency is required. Future work should validate these findings on real-world, multi-institutional CKD registries and extend the comparison to include boosting-based ensembles and deep learning approaches.

## REFERENCES

[1]. Institute for Health Metrics and Evaluation (IHME). Chronic kidney disease has more than doubled since 1990, now affecting nearly 800 million people worldwide. Global Burden of Disease Study 2023, University of Washington, 2025.
[2]. Kharoua, R. E. Chronic Kidney Disease Dataset. Kaggle, 2024.
[3]. Rubini, L. J., Soundarapandian, P., and Eswaran, P. Chronic Kidney Disease Data Set. UCI Machine Learning Repository, 2015. https://doi.org/10.24432/C5G020
[4]. Breiman, L. Random Forests. Machine Learning, 45(1), 5-32, 2001.
[5]. Quinlan, J. R. Induction of Decision Trees. Machine Learning, 1(1), 81-106, 1986.
[6]. Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research, 16, 321-357, 2002.
[7]. Powers, D. M. W. Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness and Correlation. Journal of Machine Learning Technologies, 2(1), 37-63, 2011.
[8]. Islam, M. R., and Logofătu, D. Comparative Analysis of Machine Learning Techniques for Chronic Kidney Disease Prediction Efficiency. In Engineering Applications of Neural Networks (EANN 2025), Communications in Computer and Information Science, vol. 2581, pp. 73-86, Springer, Cham, 2025.
[9]. Nguyen, D. P., Nguyen, T. T., Vu, T. T. L., Nguyen, N. N., and Nguyen, T. Q. Machine Learning Techniques in Chronic Kidney Diseases: A Comparative Study of Classification Model Performance. Clinical Kidney Journal, 2025.