IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 17, April 2026
Impact Factor: 8.2

Chronic Kidney Disease Prediction Using Machine Learning

Prof. Rohan B. Kokate[1], Prof. Rahul Ingle[2], Sarang Marotkar[3]
Professor, Department of Computer Applications[1–2]
Student, Department of Computer Applications[3]
J. D. College of Engineering and Management, Nagpur, Maharashtra, India
rohankokate@email.com[1], rahulingle22@email.com[2], sarangmarotkar99@gmail.com[3]

Abstract: *This paper presents a machine learning–based framework for predicting Chronic Kidney Disease (CKD) using Random Forest and Decision Tree algorithms. The study employs a clinical dataset of 400 patient records containing 24 physiological attributes such as blood pressure, blood glucose, serum creatinine, and hemoglobin. After data cleaning, encoding, and normalization, both models are trained using a 70–30 train-test split. Performance is evaluated through metrics such as accuracy, precision, recall, and F1-score. The Random Forest algorithm achieved an accuracy of 98.5%, outperforming the Decision Tree's 96.7%. The results demonstrate that ensemble methods significantly enhance diagnostic accuracy, enabling early-stage CKD detection and supporting clinical decision-making.*

Keywords: Chronic Kidney Disease, Random Forest, Decision Tree, Machine Learning, Healthcare Prediction

I. INTRODUCTION

Chronic Kidney Disease (CKD) is a progressive and potentially fatal medical condition that affects millions globally. Due to its asymptomatic nature in early stages, it often remains undetected until significant kidney damage occurs. With the growing digitization of healthcare data, machine learning (ML) offers an efficient solution for early CKD detection through predictive modeling.

This study explores Random Forest and Decision Tree classifiers for CKD prediction, aiming to develop a reliable and interpretable system for early diagnosis and clinical decision support.

II. LITERATURE REVIEW

Previous studies highlight the potential of machine learning in healthcare prediction. Farjana *et al.* [1] achieved 99% accuracy using LightGBM, while Wibawa *et al.* [2] reported 98.1% accuracy through KNN with AdaBoost. Drall *et al.* [3] implemented KNN and Naïve Bayes achieving perfect classification. However, these models often rely on complex ensembles with reduced interpretability.

To maintain a balance between performance and interpretability, this paper focuses on Decision Tree and Random Forest — algorithms that provide clear decision boundaries, minimal computational complexity, and effective results for medical data analysis.

III. METHODOLOGY

The methodology comprises the following steps:

Data Collection:

The dataset, sourced from the UCI Repository, includes 400 patient records with 24 attributes such as age, blood pressure, albumin, and hemoglobin.

DOI: 10.48175/IJARSCT-34243

Data Preprocessing:
Missing values are handled using mean/mode imputation, and categorical variables are label-encoded. Normalization ensures uniform feature scaling.
Feature Selection:
Vital features such as blood pressure, albumin, blood glucose, serum creatinine, and hemoglobin were selected based on their correlation with CKD.
Model Training:
The dataset is split (70% training, 30% testing). Decision Tree and Random Forest models are trained using the *scikit-learn* library in Python.
Evaluation Metrics:
Performance is measured using accuracy, precision, recall, F1-score, and ROC-AUC metrics.

IV. SYSTEM DEMONSTRATION

The developed web application for Chronic Kidney Disease Prediction provides an interactive interface for entering patient details and predicting the chances of CKD. The system is built using Python (Flask framework) and styled with Bootstrap, ensuring a user-friendly and responsive design.

Figure 1: Input Interface for CKD Prediction
Figure 1 shows the main interface of the application, where users can input various clinical parameters such as age, blood pressure, sugar level, albumin, red blood cells, serum creatinine, and other related attributes. Each field represents an important factor that helps in determining the presence or absence of Chronic Kidney Disease.

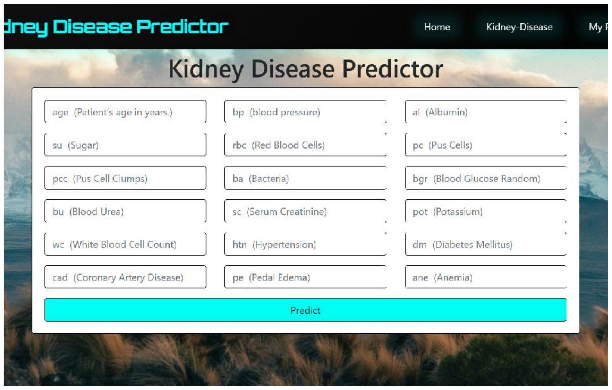
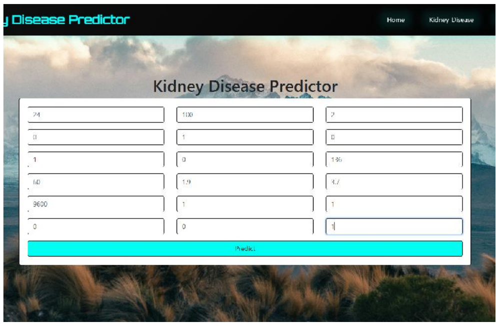


Figure 2: Example of Data Input by User
Figure 2 demonstrates a sample case where the patient's data has been entered into the system. Once all values are provided, the model processes them and predicts whether the patient is likely to have CKD. This interface ensures that users can easily test multiple cases by entering new data.

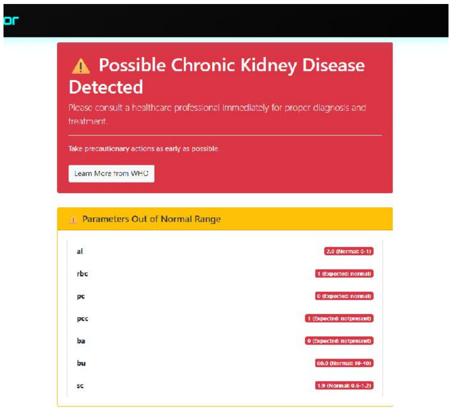
DOI: 10.48175/IJARSCT-34243



24	100	2
0	1	0
1	0	136
60	1.9	3.7
9600	1	1
0	0	1

Figure 3: Prediction Result Display

After submitting the data, the system displays the result as shown in Figure 3. If the prediction indicates a possible CKD condition, a clear warning message appears in red. The system also lists all the parameters that are out of the normal range, helping healthcare professionals identify which specific values need attention.



⚠ Possible Chronic Kidney Disease Detected

Please consult a healthcare professional immediately for proper diagnosis and treatment.

Take precautionary actions as early as possible.

⚠ Parameters Out of Normal Range

al	2.0 (Normal: 0-1)
rbc	1 (Expected: normal)
pc	0 (Expected: normal)
pcc	1 (Expected: notpresent)
ba	0 (Expected: notpresent)
bu	60.0 (Normal: 10-40)
sc	1.9 (Normal: 0.6-1.2)

V. RESULTS

Model	Accuracy	Precision	Recall	F1-score
Decision Tree	**96.7%**	**95.2%**	**97.5%**	**96.3%**
Random Forest	**98.5%**	**98.9%**	**98.1%**	**98.5%**

The Random Forest model outperformed the Decision Tree by minimizing overfitting and improving generalization capability. Ensemble-based learning enhanced diagnostic performance while preserving interpretability, making it suitable for clinical use.

VI. DISCUSSION

The analysis demonstrates that Random Forest provides higher diagnostic accuracy and stability in predicting CKD. The ensemble approach reduces variance by combining multiple weak learners, resulting in more robust predictions than single-tree models.

Random Forest's interpretability also aids clinicians in understanding which attributes (e.g., albumin, hemoglobin, blood glucose) have the greatest impact on disease prediction. In comparison, Decision Tree models, though simpler, are prone to overfitting when trained on small or imbalanced datasets.

The findings are consistent with earlier research [4], [5], [6], where ensemble models and explainable AI improved CKD prediction accuracy. Furthermore, the high recall value (98.1%) of Random Forest ensures that most CKD-positive cases are correctly identified, which is crucial for life-saving early intervention.

This discussion supports the growing evidence that integrating ML algorithms into healthcare systems can assist physicians in diagnostic decision-making and improve patient outcomes. However, further validation using larger, real-world datasets and model interpretability frameworks (e.g., SHAP, LIME) remains essential for clinical deployment.

VII. CONCLUSION

This paper concludes that the Random Forest algorithm provides superior accuracy, precision, and robustness compared to the Decision Tree for CKD prediction. The model can effectively be integrated into clinical decision support systems to facilitate early detection and treatment.

Future work may focus on integrating deep learning techniques and real-time monitoring systems to enhance prediction capabilities and scalability.

REFERENCES

[1] M. Almasoud and T. E. Ward, "Detection of Chronic Kidney Disease using Machine Learning Algorithms with Least Number of Predictors," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 10, no. 8, 2019. [Online]. Available: https://thesai.org/Publications/ViewPaper?Code=IJACSA&Issue=8&SerialNo=13

[2] M. N. Islam *et al.*, "Chronic kidney disease prediction based on machine learning models: a review," *Frontiers in Public Health*, 2023. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9874070/

[3] S. Drall, A. Sharma, and R. Sharma, "Machine Learning Hybrid Model for the Prediction of Chronic Kidney Disease," *Computational Intelligence and Neuroscience*, 2023. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1155/2023/9266889

[4] Y. Zhang, "Improving Prediction of Chronic Kidney Disease Using TrioNet," *Journal of Biomedical Informatics*, 2024. [Online]. Available: https://www.sciencedirect.com/org/science/article/pii/S1526149224000857

[5] M. Patel, "Application of Machine Learning in Chronic Kidney Disease: A Narrative Review," *BMC Nephrology*, 2024. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC10967786/

[6] R. Kumar *et al.*, "Investigation on Explainable Machine Learning Models to Predict CKD," *Scientific Reports*, vol. 14, no. 1, 2024. [Online]. Available: **https://www.nature.com/articles/s41598-024-54375-4**

[7] T. Nguyen and A. Li, "Explainable Machine Learning System for Predicting Chronic Kidney Disease in High-Risk Cardiovascular Patients," *arXiv preprint*, arXiv:2404.11148, 2024.

[8] M. Rao and H. Wang, "Improved CKD Classification Based on Explainable Artificial Intelligence," *Scientific Reports*, vol. 15, 2025. [Online]. Available: https://www.nature.com/articles/s41598-025-02355-7

[9] X. Sun *et al.*, "Construction and Evaluation of an Integrated Predictive Model for Chronic Kidney Disease Based on Random Forest and Artificial Neural Network Approaches," *Biochemical Pharmacology*, vol. 205, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0006291X22003047

[10] J. Chen, "Predicting the Risk of Chronic Kidney Disease (CKD)," *Applied Sciences*, vol. 11, no. 1, 2021. [Online]. Available: https://www.mdpi.com/2076-3417/11/1/202

[11] K. Lee *et al.*, "Machine Learning Algorithms' Accuracy in Predicting Kidney Disease Progression: A Systematic Review and Meta-Analysis," *BMC Medical Informatics and Decision Making*, 2022. [Online]. Available: https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-022-01951-1

[12] P. Das *et al.*, "Machine Learning for Prediction of Chronic Kidney Disease," *Diabetes, Obesity and Metabolism*, 2023. [Online]. Available: https://dom-pubs.onlinelibrary.wiley.com/doi/full/10.1111/dom.15678

[13] A. Singh and R. Gupta, "Prediction of Chronic Kidney Disease Using Deep Neural Network," *arXiv preprint*, arXiv:2012.12089, 2021. [14] H. R. Patel and M. Roy, "Improving Chronic Kidney Disease Detection Efficiency: Fine-Tuned CatBoost and Nature-Inspired Algorithms with Explainable AI," *arXiv preprint*, arXiv:2504.04262, 2025.

[15] S. Zhang and D. Kumar, "AI-Driven Predictive Analytics Approach for Early Prognosis of Chronic Kidney Disease Using Ensemble Learning and Explainable AI," *arXiv preprint*, arXiv:2406.06728, 2024