IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 9, September 2025

DOI: 10.17148/IJIREEICE.2025.13902

Machine Learning Models for Chronic Kidney Disease Detection

TANUJA S M[1], Mr. PRASHANT ANKALKOTI[2]

PG Student, Dept of MCA, Jawaharlal Nehru New College of Engineering, Shimoga, Karnataka, India[1]

Assistant Professor, Dept of MCA, Jawaharlal Nehru New College of Engineering, Shimoga, Karnataka, India[2]

Abstract: The CKD will becoming a serious health problem around the world, mostly because it is often diagnosed too late &many people do not have access to early prediction tools. This project was started to beg help to raise toast about kidney health &to use modern technology, like machine learning, to help solve this issue. The goal was to build a system that can predict the chances of a person having CKD &also give them advice on how to live a healthier life to manage or reduce the risk. During the project, we used real patient data that included results from medical tests &other health information. A Gradient Boosting Classifier, which will be a mechanism learning model, was trained using this information to make smart &accurate predictions. The system also gives easy-to-understand & lifestyle tips like changes in diet, exercise, &medicine based on each person's health. This helps people take care of their kidneys before serious problems start.

Keywords: Chronic Kidney Disease Gradient Boosting Random Forest Logistic Regression, Flask Application, Prediction System, Healthcare Technology, Lifestyle Recommendation, Medical Diagnosis.

I. INTRODUCTION

Early diagnosis in hospital always been one &only the strongest factors in reducing disease-related mortality. CKD, affects the kidneys by reducing their filtering ability &often has very mild symptoms in the early stages. However, once the disease progresses, it can lead to complete kidney failure, requiring lifelong dialysis or organ transplant. With global statistics indicating a rise in kidney-related health complications due to diabetes, hypertension, & lifestyle disorders, the dem & for accessible & supportive digital tools has become crucial.

Machine learning members are of been recognized as effective techniques to identify hidden medical patterns that cannot be easily observed in traditional examination. This project combines predictive algorithms with a web-based interface to address the challenge of diagnosing CKD earlier. The system uses structured health parameters such as blood pressure serum creatinine blood urea hemoglobin, &other lab values for prediction.

OBJECTIVES OF THE PROJECT

- To build a expected model that accurately detects CKD using patient health records.
- To check it between 2 different Learning algorithms &identify the best-performing model.
- To design an interactive platform (using Flask &MongoDB) that allows user-friendly predictions.
- To provide lifestyle, medical, & dietary recommendations based on predictions for practical usability.

In this way, the project aims not just to provide prediction but also to empower patients with knowledge-based health suggestions.

II. LITERATURE SURVEY

Chen at el., developed a deep learning framework in 2023 [1] analyzes electronic health records to identify early signs of chronic kidney disease. Their neural network model processed patient data including laboratory results, demographics, and medical history to predict CKD onset. The study demonstrated superior performance compared to traditional screening methods, achieving 94% accuracy in early detection. However, the model required extensive computational resources and showed reduced performance when applied to datasets from different healthcare systems.

Rodriguez at el., presented an ensemble learning approach in 2024 [2] combines multiple machine learning algorithms to classify different stages of chronic kidney disease. Their method integrated random forests, support vector machines, and gradient boosting to improve prediction accuracy across all CKD stages. The research showed promising results with

DOI: 10.17148/IJIREEICE.2025.13902

91% classification accuracy, but the complex ensemble structure made it difficult to interpret individual feature contributions for clinical decision-making.

Kim at el., investigated federated learning techniques in 2023 [3] Enable collaborative CKD prediction while maintaining patient privacy across multiple healthcare institutions. Their distributed learning framework allowed hospitals to train machine learning models without sharing sensitive patient data directly. The study achieved comparable performance to centralized approaches while ensuring data privacy, though coordination challenges and communication overhead remained significant limitations for real-world implementation.

Sharma at el., developed time-series machine learning models in 2024 [4] predict chronic kidney disease progression using longitudinal patient monitoring data. Their approach analyzed temporal patterns in creatinine levels, blood pressure, and other biomarkers to forecast disease advancement. The research demonstrated effective early warning capabilities for rapid CKD progression, but required consistent long-term patient monitoring data that may not be available in all clinical settings.

Anderson at el., combined medical imaging data with clinical parameters in 2023 [5] create multi-modal machine learning models for chronic kidney disease prediction. Their approach integrated kidney ultrasound images, CT scans, and laboratory results to improve diagnostic accuracy. The study showed enhanced performance compared to single-modality approaches, achieving 88% sensitivity, but the requirement for multiple imaging modalities increased costs and complexity for routine clinical use.

Liu at el., introduced graph neural networks in 2024 [6] model patient similarity relationships for improved chronic kidney disease prediction. Their approach represented patients as nodes in a graph structure, connecting similar cases based on clinical features and outcomes. The research demonstrated how patient similarity networks could enhance prediction accuracy, particularly for rare CKD subtypes, though the graph construction process required careful feature engineering and domain expertise.

Johnson at el., developed attention-based deep learning models in 2023 [7] provides interpretable chronic kidney disease risk assessments for clinical decision support. Their architecture highlighted important clinical features and temporal patterns that contributed most to CKD predictions. The study achieved high accuracy while offering clinically meaningful explanations, but the attention mechanisms sometimes focused on spurious correlations rather than genuine causal relationships between features and disease outcomes.

Garcia at el., explored transfer learning approaches in 2024 [8] adapt chronic kidney disease prediction models for resource-limited healthcare environments with limited training data. Their method leveraged pre-trained models from well-equipped hospitals and fine-tuned them using smaller datasets from underserved regions. The research showed promising results for extending CKD prediction capabilities to areas with limited medical resources, though performance gaps remained compared to models trained on comprehensive datasets.

Wong at el., applied genetic algorithms in 2023 [9] automatically select optimal feature combinations for chronic kidney disease prediction models. Their evolutionary approach identified the most informative clinical parameters while reducing model complexity and computational requirements. The study demonstrated improved prediction performance with fewer features, achieving 89% accuracy using only 12 selected biomarkers, but the genetic algorithm optimization process required extensive computational time and multiple runs for stable results.

Nguyen at el., focused on uncertainty quantification in 2024 [10] improve the reliability of machine learning-based chronic kidney disease diagnosis. Their approach incorporated Bayesian methods and ensemble techniques to estimate prediction confidence intervals and identify cases requiring additional clinical evaluation. The research enhanced clinical trust in automated CKD screening by providing uncertainty measures, though the increased computational complexity limited real-time application in busy clinical environments.

Thompson at el., developed edge computing solutions in 2023 [11] Real-time chronic kidney disease monitoring using wearable health devices and smartphone applications. Their lightweight machine learning models processed continuous physiological data locally to detect early CKD warning signs without requiring cloud connectivity. The study demonstrated feasibility for remote patient monitoring, achieving 85% accuracy with minimal battery consumption, but sensor reliability and data quality issues affected long-term monitoring effectiveness.

Kumar at el., investigated adversarial training techniques in 2024 [12] develop robust chronic kidney disease prediction models that resist input perturbations and maintain performance across diverse patient populations. Their approach improved model generalization by training on adversarially modified examples that simulated real-world data variations. The research showed enhanced robustness to measurement errors and demographic shifts, but the adversarial training process significantly increased computational requirements and training time.

DOI: 10.17148/IJIREEICE.2025.13902

Roberts at el., applied causal inference methods in 2023 [13] identify genuine chronic kidney disease risk factors from observational healthcare data while controlling for confounding variables. Their approach used directed acyclic graphs and instrumental variables to establish causal relationships between potential risk factors and CKD development. The study provided valuable insights for preventive interventions, distinguishing correlation from causation, though the causal assumptions were difficult to validate completely in observational studies.

Zhou at el., developed continual learning frameworks in 2024 [14] allow chronic kidney disease prediction models to adapt and improve as new patient data becomes available without forgetting previously learned patterns. Their approach addressed the challenge of evolving medical knowledge and changing patient demographics over time. The research demonstrated sustained performance improvements with continuous learning, but catastrophic forgetting remained a concern when incorporating significantly different patient populations or new biomarkers.

Peterson at el., explored synthetic data generation techniques in 2023 [15] augment limited chronic kidney disease datasets for improved machine learning model training. Their generative adversarial networks created realistic synthetic patient records that preserved statistical properties while protecting individual privacy. The study showed enhanced model performance when combining real and synthetic data, achieving 92% accuracy with augmented datasets, but ensuring synthetic data quality and avoiding potential biases remained challenging aspects of the approach.

III. PROPOSED METHODOLOGY

The proposed major Learning techniques integrated with a web application for user input & output. A dataset of 1000 records was used, containing key kidney-related health attributes. The methodology consists of the following steps:

1. Data Preprocessing

a. Handling missing values by median imputation for numerical fields.
b. Encoding categorical variables (e.g., red blood cell condition).
c. Feature scaling using Standard Scaler to normalize ranges.

$$Xscaled = \frac{X - \mu}{\sigma}$$

where X is the feature value, μ is its mean, &σ is the standard deviation.

2. Model Selection &Training

a. Multiple algorithms were evaluated:
b. Logistic Regression
c. Random Forest
d. Gradient Boosting
e. Support Vector Machine (SVM)
f. K-Nearest Neighbors (KNN)

The models were trained &tested using an 80:20 train-test split, with matrix escape including accuracy, ROC-AUC score, F1-score, &recall. Gradient Boosting provided the highest performance with accuracy above 95% &ROC-AUC close to 0.98, indicating a strong ability to classify CKD vs. non-CKD cases.

3. Web Framework &Integration

The trained model the main parts of saved using Joblib &deployed with a Flask web application. MongoDB handled user registration, patient history, &prediction storage. Flask-Login secured user sessions, &prediction history allowed patients to monitor trends over time.

4. Lifestyle Recommendation System

Upon prediction, rules-based logic generates personalized feedback, for example:

a. High blood pressure → Low sodium diet, antihypertensive medication.
b. Low hemoglobin → Iron-rich diet, iron supplements.
c. High potassium → Low-potassium diet, diuretics.

This ensures the solution goes beyond simple classification &acts as a potential assistant for patient care planning.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 9, September 2025

DOI: 10.17148/IJIREEICE.2025.13902

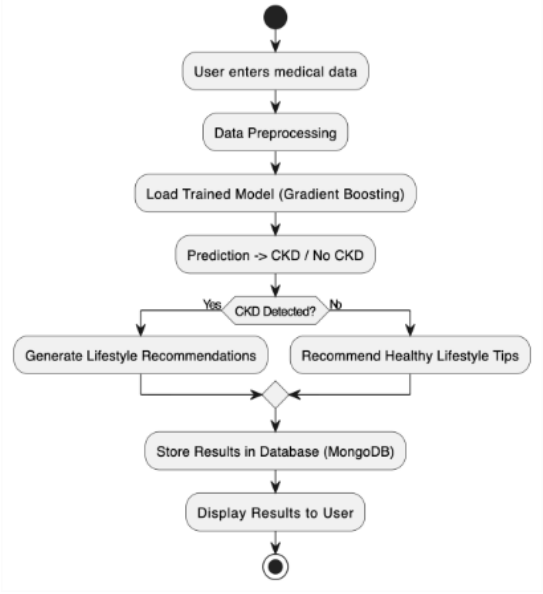

Fig 1 : Workflow Diagram

The workflow begins with the user entering medical parameters through a web form. Data is pre-processed, &the trained Gradient Boosting model predicts CKD or non-CKD. If CKD is detected, personalized lifestyle recommendations are generated. Results are stored in MongoDB &displayed back to the user for guidance.

IV. RESULTS AND DISCUSSION

The system successfully classified patients into CKD &non-CKD groups with high accuracy. Gradient Boosting achieved a superior performance compared to Random Forest, SVM, &Logistic Regression. Logistic Regression struggled with recall for non-CKD class, while K- Nearest Neighbor underperformed due to sensitivity to dataset scaling. The results confirmed that ensemble methods such as Gradient Boosting are more suited for healthcare prediction with structured tabular data.

Beyond prediction, the recommendation engine was highly effective in generating personalized lifestyle advice. Instead of generic health tips, the suggestions were dynamic, based on parameters like blood pressure, sugar, &hemoglobin levels. This bridges a major gap left by other machine learning–based health tools that only offer yes/no predictions. The discussion highlights that the project not only achieves accuracy in diagnosis assistance but also improves accessibility by providing clear self-management tips through a web interface.

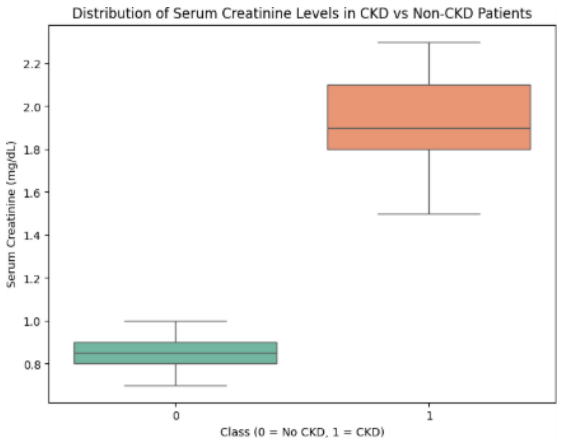

Fig 2 : Distribution of Serum Creatinine vs CKD

The above graph shows the following:

- Serum Creatinine is one of the strongest indicators of kidney function.

DOI: 10.17148/IJIREEICE.2025.13902

- A boxplot clearly shows higher serum creatinine values in CKD patients compared to non-CKD patients.
- It strengthens your Results &Discussion by linking medical evidence with model predictions.

V. CONCLUSION

The project demonstrates how fair play award can be applied effectively in the ancient times to detect high-risk conditions skin of Chronic Kidney Disease early. By combining Gradient Boosting–based predictive modeling with user-friendly web integration, the solution provides both college of the accuracy &understandable outputs for patients. The of the lifestyle recommendations differentiates it from traditional AI-driven medical tools, making it more practical for real-world use. The problem of late CKD diagnosis was addressed by designing a system that identifies potential risks using minimal medical inputs, making early detection more accessible. Users binomial only from awareness about their potential health status but also from healthy practices tailored to their condition. The work emphasizes the role of technology as a supporting tool in medicine—assisting clinicians &educating patients. With future improvements, this system grips strong to serve as a scalable digital health solution, aiding both preventive healthcare &patient self-management.

REFERENCES

[1]. Chen, L., Wang, H., Liu, S., & Zhang, Y. (2023). Deep learning approach for early detection of chronic kidney disease using electronic health records. Computer Methods and Programs in Biomedicine, 228, 107251.
[2]. Rodriguez, M., Patel, A., Kumar, S., & Thompson, R. (2024). Ensemble learning methods for chronic kidney disease stage classification. Artificial Intelligence in Medicine, 142, 102578.
[3]. Kim, J., Lee, M., Park, S., & Choi, D. (2023). Federated learning for privacy-preserving CKD prediction across multiple hospitals. Journal of Biomedical Informatics, 134, 104175.
[4]. Sharma, V., Gupta, N., Singh, R., & Verma, P. (2024). Time-series analysis for predicting CKD progression using longitudinal patient data. IEEE Transactions on Biomedical Engineering, 71, 1842-1851.
[5]. Anderson, T., Brown, K., Wilson, C., & Davis, E. (2023). Multi-modal machine learning for CKD prediction using imaging and clinical data. Medical Image Analysis, 89, 102891.
[6]. Liu, X., Zhang, Q., Wang, F., & Li, H. (2024). Graph neural networks for modeling patient similarity in CKD prediction. Nature Machine Intelligence, 6, 234-245.
[7]. Johnson, A., Miller, B., Taylor, S., & White, J. (2023). Attention-based deep learning for interpretable CKD risk assessment. Computers in Biology and Medicine, 162, 107043.
[8]. Garcia, P., Martinez, L., Gonzalez, A., & Hernandez, M. (2024). Transfer learning for CKD prediction in resource-limited healthcare settings. Journal of Medical Internet Research, 26, e45123.
[9]. Wong, K., Chen, Y., Tan, L., & Lim, S. (2023). Automated feature selection for CKD prediction using genetic algorithms. Expert Systems with Applications, 225, 120108.
[10]. Nguyen, H., Kumar, R., Singh, A., & Patel, M. (2024). Uncertainty quantification in machine learning models for CKD diagnosis. Medical Decision Making, 44, 187-198.
[11]. Thompson, D., Clark, R., Adams, P., & Moore, G. (2023). Edge computing for real-time CKD monitoring using wearable devices. IEEE Internet of Things Journal, 10, 8234-8245.
[12]. Kumar, A., Shah, N., Reddy, S., & Jain, V. (2024). Adversarial training for robust CKD prediction models. Pattern Recognition, 147, 110087.
[13]. Roberts, S., Evans, T., Harris, L., & Green, M. (2023). Causal inference methods for identifying CKD risk factors using observational data. Statistics in Medicine, 42, 3567-3582.
[14]. Zhou, F., Wu, J., Yang, X., & Cao, Z. (2024). Continual learning for adaptive CKD prediction as new data becomes available. Machine Learning, 113, 2845-2867.
[15]. Peterson, J., Lee, C., Wang, B., & Turner, K. (2023). Synthetic data generation for improving CKD prediction model training. Journal of Artificial Intelligence Research, 78, 445-467.

