iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: XI Month of publication: November 2025

DOI: https://doi.org/10.22214/ijraset.2025.75630

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XI Nov 2025- Available at www.ijraset.com

Chronic Kidney Disease Detection Using Machine Learning Model

Vadthyavath Shirisha[1], Dr. M. Dhanalakshmi[2]
[1]M. Tech, Data Science, Department of Information Technology, Jawaharlal Nehru Technological University Hyderabad, UCESTH, India
[2]Professor of IT Dept & Deputy Director of DILT, Department of Information Technology, Jawaharlal Nehru Technological University Hyderabad, UCESTH, India

Abstract: Chronic Kidney Disease (CKD) is a long-term medical condition in which the kidneys gradually lose their ability to function properly. Also known as chronic renal failure, this disease progresses over a period of several months or even years, often without obvious symptoms in its early stages. Because of this, CKD is frequently diagnosed only through routine medical examinations, especially in individuals with existing risk factors like diabetes, high blood pressure, or a family history of kidney issues. Detecting CKD early is essential for managing the illness and preventing serious complications. With early diagnosis being difficult due to the lack of visible signs, healthcare professionals have increasingly turned to Machine Learning (ML) as a modern solution. ML techniques can analyze clinical datasets to detect patterns and predict disease risk with impressive accuracy. Among these techniques, the Random Forest algorithm has proven particularly effective in identifying CKD, outperforming many other models in terms of prediction reliability.
Keywords: Machine Learning, Disease Detection, Random Forest, Decision Tree.

I. INTRODUCTION

Chronic Kidney Disease (CKD) is a long-term health disorder marked by the gradual deterioration of kidney function. The kidneys play a central role in filtering waste, maintaining fluid and electrolyte balance, supporting red blood cell production, managing blood pressure, and regulating essential minerals like calcium, sodium, and potassium. When the kidneys begin to lose their ability to perform these tasks, various complications can arise. The most common contributors to CKD include chronic conditions such as diabetes and hypertension. Other causes may involve infections, autoimmune diseases, or the prolonged use of harmful medications. In its early stages, CKD is usually asymptomatic, making early identification difficult without routine medical testing. As the condition progresses, patients may develop fatigue, swelling, changes in urination, nausea, or cardiovascular issues. Left untreated, CKD may lead to end-stage renal disease (ESRD), which requires life-saving treatments such as dialysis or organ transplantation. CKD is typically classified into five progressive stages, with stage 1 reflecting minor impairment and stage 5 representing complete kidney failure. Accurate staging depends on the glomerular filtration rate (GFR), a clinical metric used to assess how well the kidneys are filtering blood. Recognizing the disease early is key to improving outcomes, yet early detection remains a major hurdle in clinical practice. To overcome this, recent advancements in data science have introduced the application of machine learning (ML) in healthcare. ML algorithms are capable of analyzing large sets of patient data to uncover trends and predict disease risk with considerable accuracy.

II. RELATED WORK

Chronic Kidney Disease (CKD) has been extensively studied in recent years due to its growing prevalence and impact on global health. Prior research has focused on identifying key risk factors such as diabetes, hypertension, obesity, and genetic predisposition, as well as developing early detection methods through biochemical markers like serum creatinine, glomerular filtration rate (GFR), and urine albumin levels. Several studies have explored the use of machine learning algorithms for predicting CKD progression, including decision trees, support vector machines, and neural networks, demonstrating significant accuracy in early diagnosis. Additionally, work has been done on improving patient management through predictive analytics, lifestyle interventions, and personalized treatment plans. International health organizations and national health systems have also implemented screening programs and awareness campaigns to reduce late-stage CKD diagnosis and improve patient outcomes. This existing body of work provides a strong foundation for developing more accurate, automated, and cost-effective CKD prediction and management systems.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XI Nov 2025- Available at www.ijraset.com

III. METHODOLOGY

This section explains the methodology used to develop a machine learning-based system for predicting Chronic Kidney Disease (CKD) using clinical data. The primary goal is to build an accurate, structured, and accessible model that can assist in early detection of CKD through a web-based interface. The process begins by collecting the CKD dataset from the UCI Machine Learning Repository. The data is then cleaned and preprocessed to handle missing values, normalize features, and encode categorical variables. The Random Forest algorithm is selected for its ability to handle complex datasets, reduce over fitting, and deliver high prediction accuracy. After the model is trained and validated, it is evaluated using standard metrics such as accuracy, precision, recall, and confusion matrix to assess its performance. Once satisfactory results are achieved, the trained model is integrated into a Flask-based web application. This interface allows users to enter patient information through a simple form and receive real-time predictions about CKD status and severity level (e.g., high or low). The methodology ensures that every stage from data preparation to deployment is logically structured, reproducible, and tailored for healthcare applications, with the aim of supporting early diagnosis and improving patient outcomes. The design framework of this system outlines how different components work together to achieve accurate predictions. It serves as a visual and conceptual guide for understanding the key modules, their responsibilities, and how information moves through the system. A clearly defined architecture is especially useful at the beginning of development, as it improves both clarity and planning efficiency. In this application, users begin by entering specific health related values such as diagnostic and clinical indicators. These inputs are first sent through a preprocessing pipeline, where they are cleaned and converted into a format suitable for machine learning tasks. The refined data is then divided into two parts: one set is used to train the algorithm, while the other is reserved to test its performance. The system uses the Random Forest technique to create a predictive model. This involves generating a collection of decision trees, where each tree gives an individual output. The final decision is made based on the majority of votes from these trees. After the model is trained, it is tested using the reserved dataset to evaluate its effectiveness. Once predictions are generated, results are shown to the user and can optionally be saved for future insights or reporting.

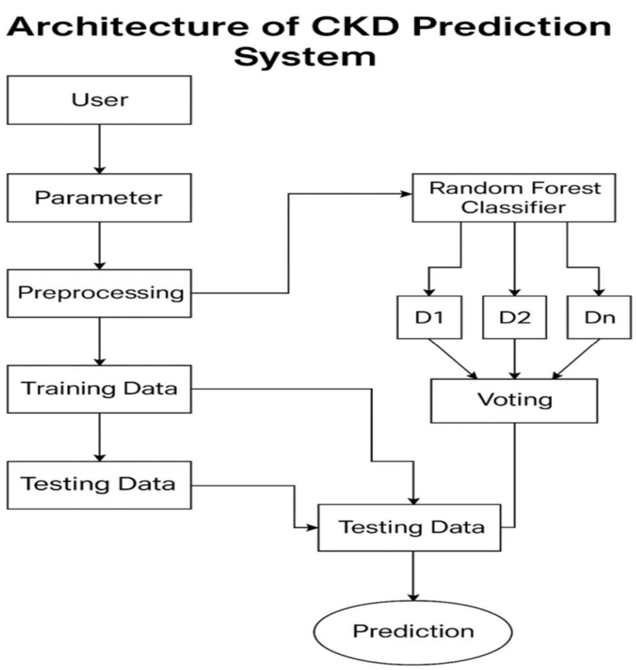


Fig1. Architecture Design

A. Modules

1) Data Collection and Preprocessing

The dataset was collected from the UCI Machine Learning Repository, containing 400 patient records and 26 features related to health parameters (e.g., blood pressure, red blood cells, haemoglobin, serum creatinine).

Data pre-processing included handling missing values, and splitting into training and testing sets using train test split.

2) Model Selection and Training

The Random Forest algorithm was selected due to its strong performance with clinical datasets and its effectiveness in minimizing the risk of over fitting.

This technique follows an ensemble learning approach, where numerous decision trees are generated during training. The final output is determined based on the majority voting among the individual trees.

The model development and training processes were carried out using the Scikit-learn library in Python, which offers robust tools for machine learning implementation.

3) Model Evaluation

Once the training phase was completed, the model's performance was assessed using a separate testing dataset.

Key evaluation metrics included Accuracy, Precision, Recall, F1-Score, and the Confusion Matrix.

These metrics provided a comprehensive view of how well the model was able to detect Chronic Kidney Disease, helping to evaluate its reliability and predictive capability.

4) Model Saving

The trained Random Forest model was saved using the pickle library so it could be reused without retraining every time.

5) Web Interface Development

A simple and interactive web application was developed using Flask.

Users can enter clinical values into the form and the model will instantly predict whether the person is likely to have CKD.

The system provides results in a clear and understandable format.

6) Integrated Prediction System

7) All components pre-processing, trained model, and user interface, were integrated into a single system. The goal is to make CKD prediction faster, more *accurate*, and accessible to healthcare .

IV. RESULTS AND EVALUATIONS

In the below screen, it shows the accuracy score of different machine learning models used for predicting Chronic Kidney Disease (CKD). The scores represent how well each model performed during testing.

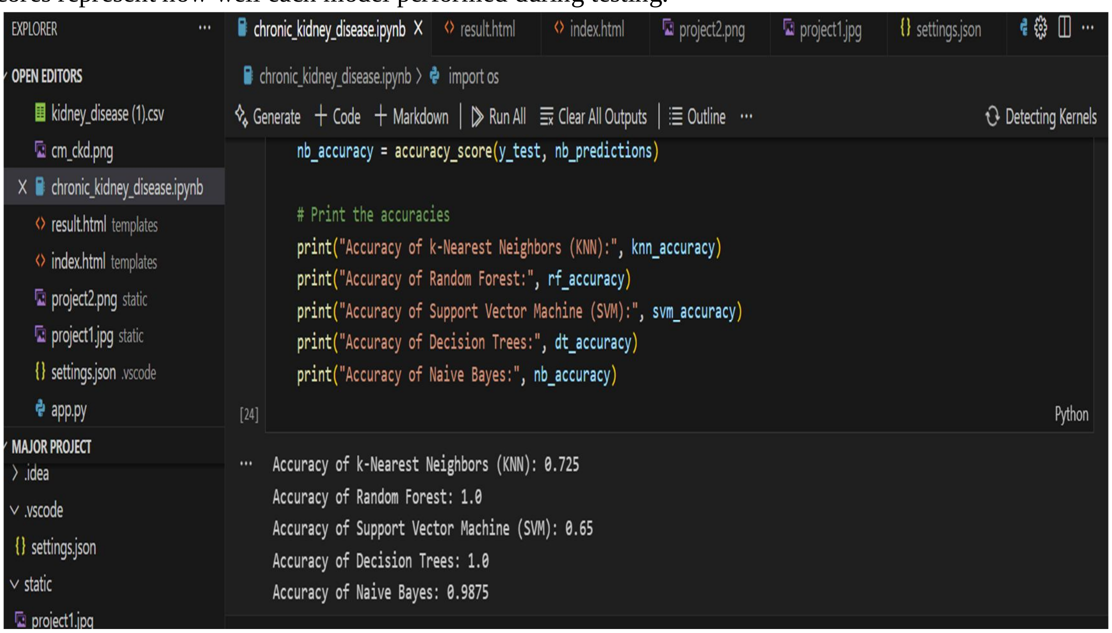


Fig : Accuracy of Different Classifiers on CKD Dataset

A. Output Screens

Distribution of Attributes

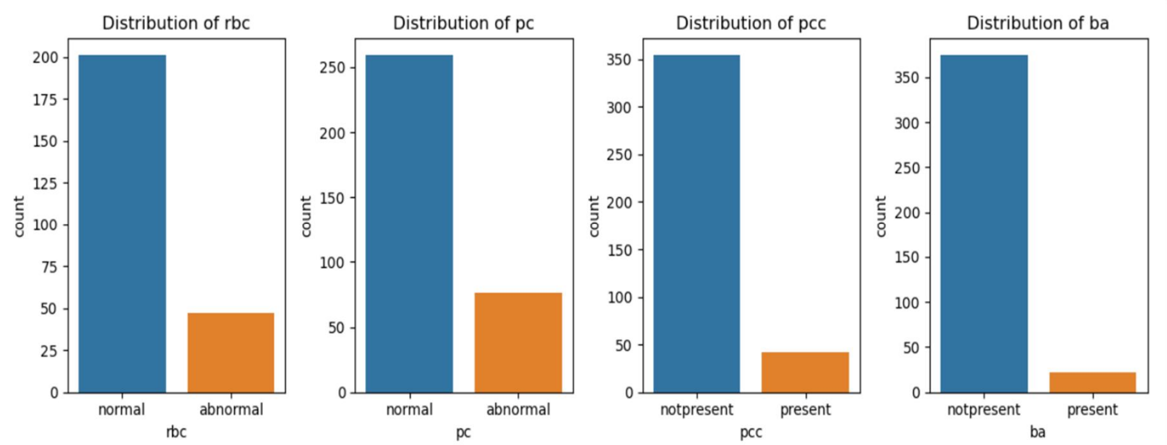


Fig: Distribution of RBC, PC, PCC, and BA Attributes

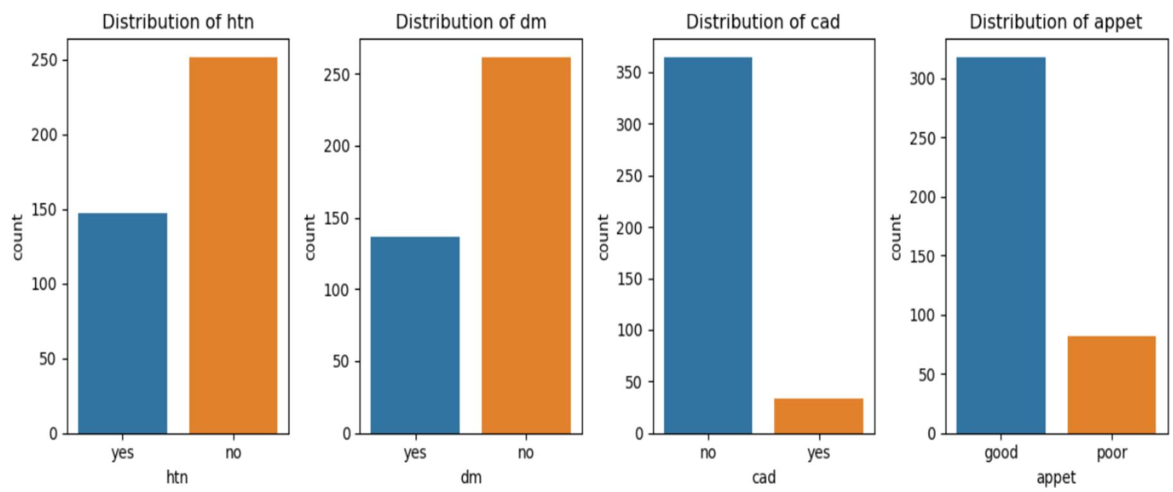


Fig: Distribution of HTN, DM, CAD, and Appet Attributes

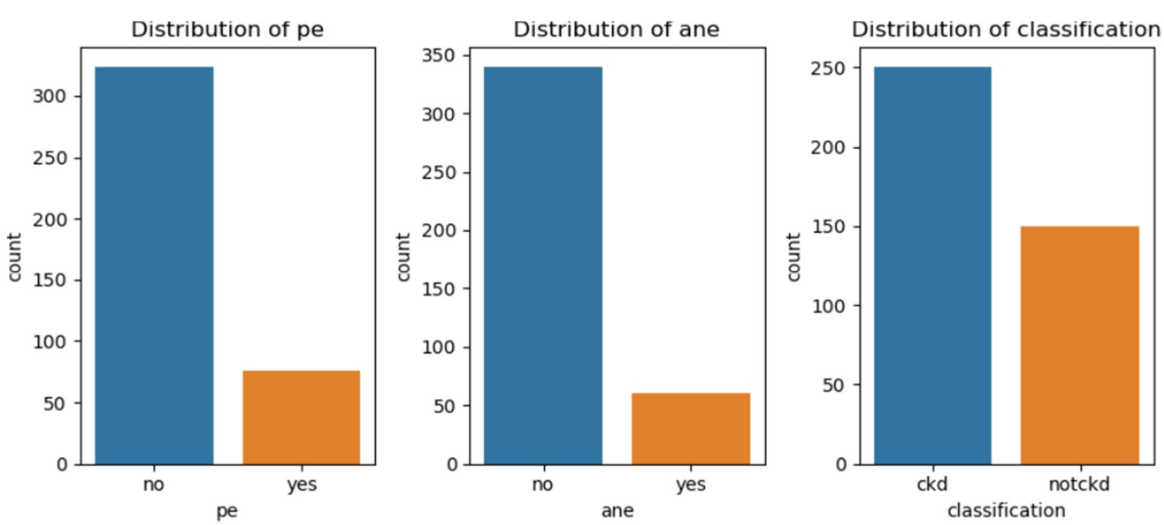
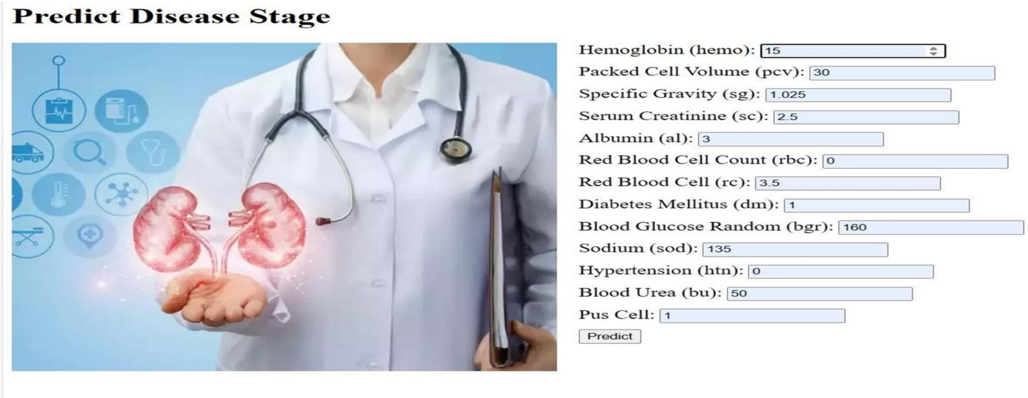


Fig: Distribution of PE, Ane, and Classification (CKD vs Not CKD)

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XI Nov 2025- Available at www.ijraset.com

B. Final Output



This figure shows the user interface of the CKD prediction system. Users input their medical parameters such as hemoglobin level, specific gravity, serum creatinine, and others into the form. After clicking the “Predict” button, the system processes the input and predicts whether the individual is suffering from Chronic Kidney Disease and its severity level.

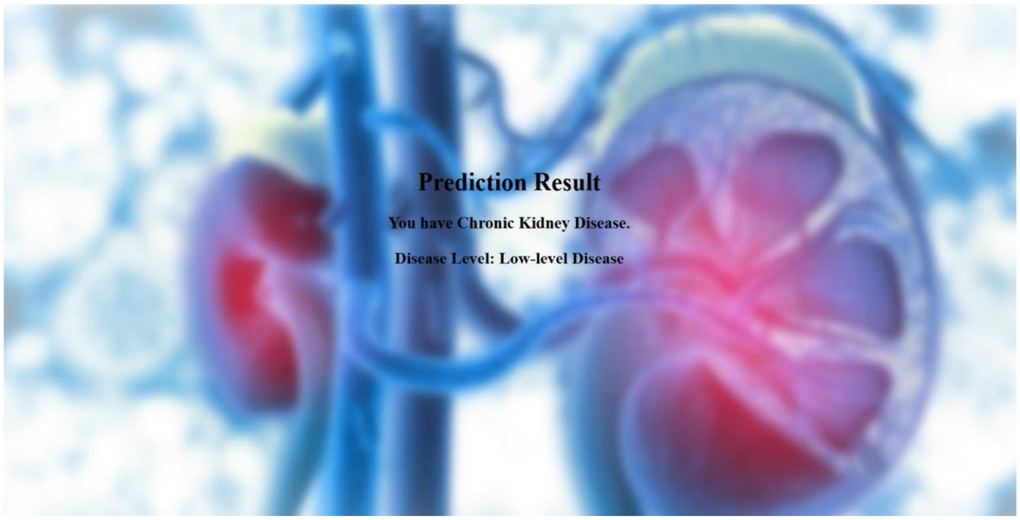


Fig: Prediction Result for Chronic Kidney Disease showing "Low-level Disease" classification.

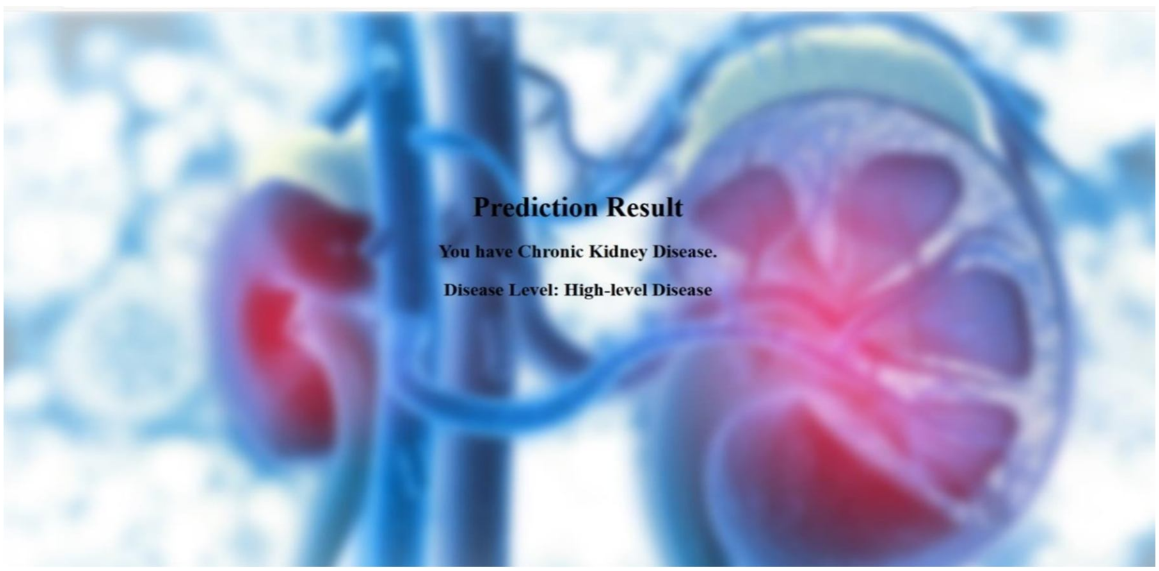


Fig: Prediction Result for Chronic Kidney Disease showing "High-level Disease" classification.

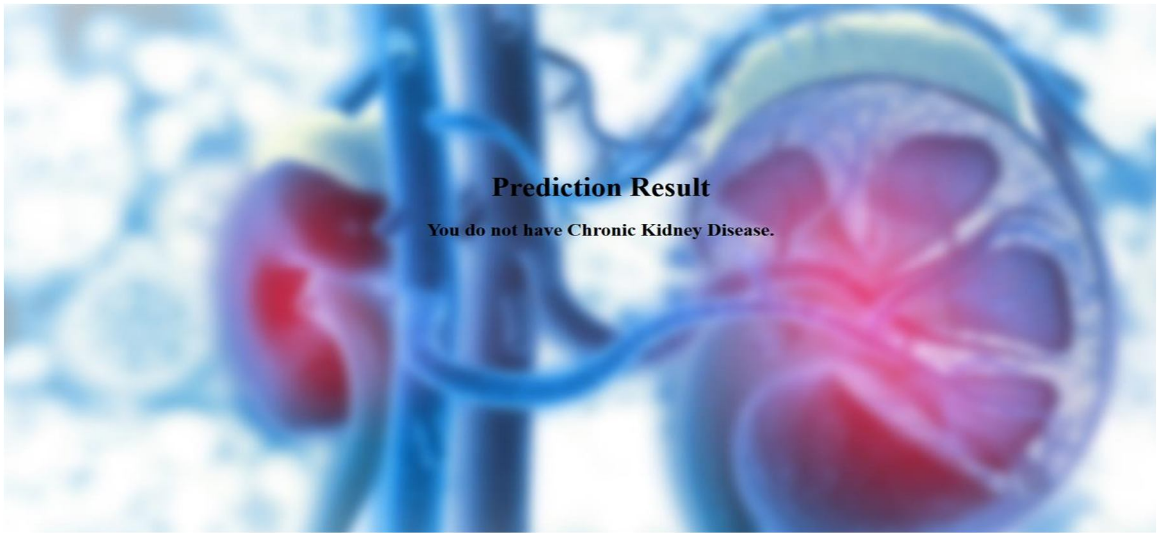


Fig: Prediction Result for Chronic Kidney Disease showing "No Disease" classification.

The results of this CKD prediction project were analyzed based on the performance of various machine learning models applied to structured medical datasets after pre-processing and data cleaning. The evaluation focused on key classification metrics including accuracy, precision, recall, F1-score, and confusion matrix values for each of the models.

Initially, individual models such as Naive Bayes, k-Nearest Neighbors (KNN), Decision Tree, Random Forest, and Support Vector Machine (SVM) were trained on the CKD dataset. The features included laboratory test values and medical history indicators (e.g., haemoglobin, albumin, diabetes, blood urea, etc.).

V. CONCLUSION

In this project, a machine learning-based system was developed to predict Chronic Kidney Disease (CKD) using the Random Forest algorithm. The model was trained on a publicly available dataset and successfully classified individuals as either healthy or affected by CKD, along with indicating the severity level (high or low). The Random Forest classifier demonstrated strong performance in terms of accuracy and reliability, making it a suitable choice for medical prediction tasks. The system's ability to deliver clear, real-time results through a user-friendly web interface enhances its usability in practical healthcare settings. By enabling early detection and classification of CKD, this model can assist medical professionals in making informed clinical decisions and improve patient outcomes. The future scope for a chronic kidney disease (CKD) project using a random forest algorithm could involve refining the model with more data, enhancing feature selection techniques, exploring ensemble methods, and integrating it into healthcare systems for early detection and personalized treatment recommendations. Another potential avenue could be the development of a user-friendly mobile app that integrates the CKD prediction model, allowing users to assess their risk of CKD and receive personalized recommendations for lifestyle modifications.

REFERENCES

[1] M. P. N. M. Wickramasinghe, D. M. Perera, and K. A. D. C. P. Kahandawaarachchi, "Dietary prediction for patients with chronic kidney disease (CKD) by considering blood potassium level using machine learning algorithms," 2017 IEEE Life Sciences Conference (LSC), Sydney, NSW, 2017, pp. 300–303.

[2] U. N. Dulhare and M. Ayesha, "Extraction of action rules for chronic kidney disease using Naïve Bayes classifier," 2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Chennai, 2016, pp. 1–5.

[3] R. Devika, S. V. Avilala, and V. Subramaniyaswamy, "Comparative Study of Classifier for Chronic Kidney Disease Prediction Using Naïve Bayes, KNN and Random Forest2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2019, pp. 679–684.

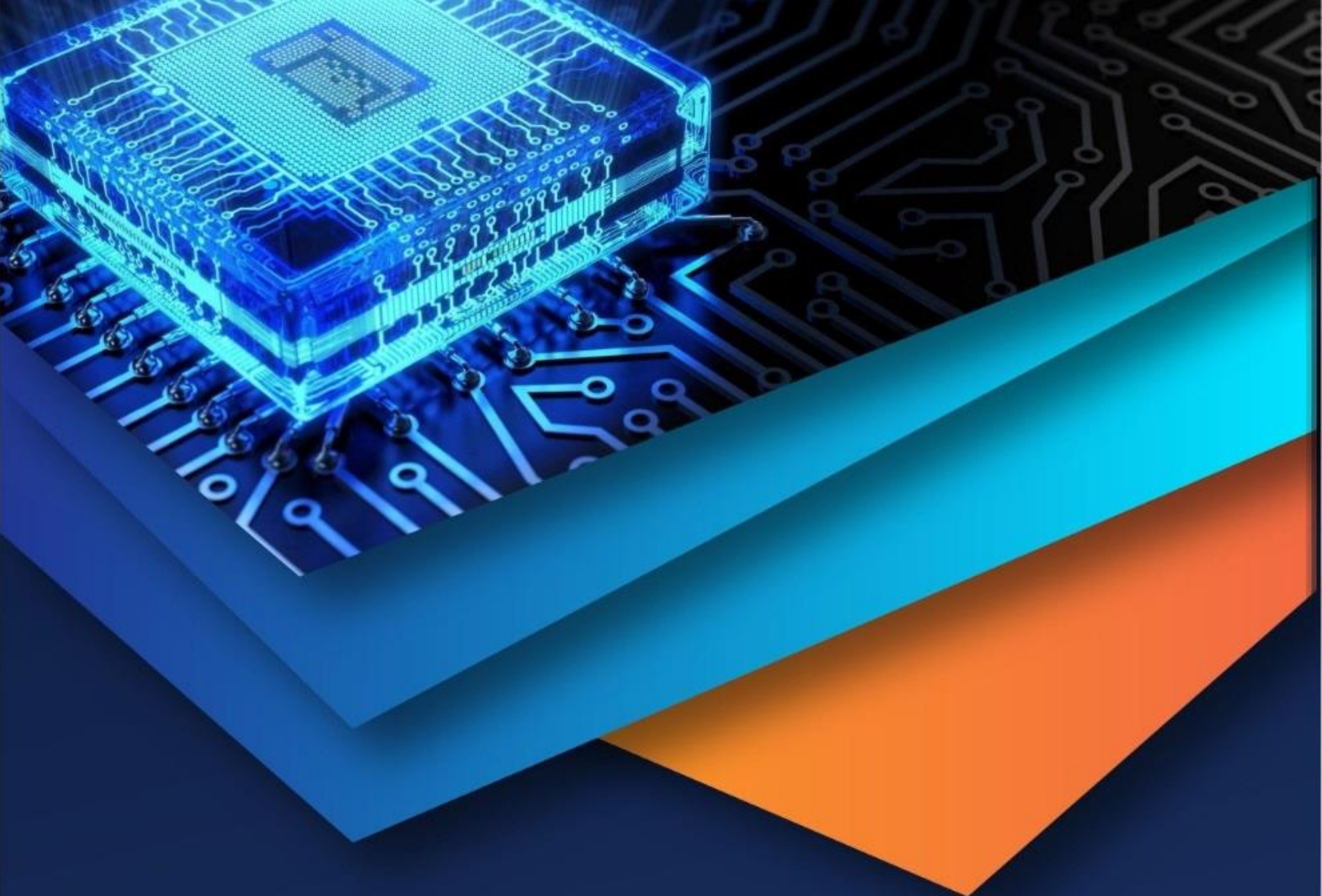
[4] G. Kaur and A. Sharma,"Predict chronic kidney disease using data mining algorithms in Hadoop," 2017 International Conference on Inventive Computing and Informatics (ICICI), Coimbatore, 2017, pp. 973–979.

[5] Arif-Ul-Islam and S. H. Ripon, "Rule Induction and Prediction of Chronic Kidney Disease Using Boosting Classifiers, Ant-Miner and J48 Decision Tree," 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), Casazza, Bangladesh, 2019, pp. 1–6

[6] 6. S. Vijayarani and S. Dayanand, "Kidney Disease Prediction Using SVM and ANN Algorithms," International Journal of Computing and Business Research (IJCBR), vol. 6, no. 2, 2015.

[7] D. Sisodia and D. S. Sisodia, "Prediction of Diabetes Using Three Classification Algorithms," International Conference on Computational Intelligence and Data Science (ICCIDS), 2018.

[8] M. S. Gharibdousti, K. Azimi, S. H. Kal, and D. H. Won, "Prediction of Chronic Kidney Disease Using Data Mining Techniques," 2018.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

TOGETHER WE REACH THE GOAL

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY