IJITCE

ISSN 2347- 3657

International Journal of Information Technology & Computer Engineering

Email : ijitce.editor@gmail.com or editor@ijitce.com

International Journal of Information Technology & Computer Engineering

ISSN 2347-3657

Volume 13, Issue 2, 2025

Chronic kidney disease prediction based on machine learning algorithms

V. SRILAKSHMI[1],Dr.K.CHAITANYA[2],S.PAVANI[3],J.SAI CHAKRA DHAR RAO[4],J.KRISHNA CHAITANYA [5]

Assistant Professor[2] , Department of Computer Science & Engineering,University College of Engineering and Technology,Acharya Nagarjuna University,Nagarjuna Nagar,Guntur,AP, INDIA
Email:anu.konda.chaitanya@gmail.com[2]

UG Students[1,3,4,5], Department of Computer Science & Engineering,University College of Engineering and Technology, Acharya Nagarjuna University, Nagarjuna Nagar,Guntur,AP, INDIA

Email:vsrilakshmi498@gmail.com[1],pavanisunkari912@gmail.com[3],saichakradharrao2004@gmail.com[4],krishnachaitanyajanga04@gmail.com[5]

Abstract: Chronic kidney disease (CKD) is a dangerous ailment that can last a person's entire life and is caused by either kidney malignancy or decreased kidney function. It is feasible to halt or slow the progression of this chronic disease to an end-stage wherein dialysis or surgical intervention is the only method to preserve a patient's life. Earlier detection and appropriate therapy can increase the likelihood of this happening. Throughout this research, the potential of several different machine learning approaches for providing an early diagnosis of CKD has been investigated. In recent years, machine learning (ML) algorithms have become a powerful tool in medical diagnosis, offering the potential to predict kidney disease with high accuracy. This paper investigates the application of various ML techniques, including Decision Trees, Support Vector Machines (SVM), Random Forests, Gradient Boost Classifier, Xgboost, KNN, and Logistic Regression, in predicting kidney disease using clinical data. Therefore, in our approach, we investigate the link that exists between data factors as well as the characteristics of the target class. We are capable of constructing a collection of prediction models with the help of machine learning and predictive analytics. The performance of these algorithms is evaluated based on metrics such as accuracy, sensitivity, specificity, and receiver operating characteristic (ROC).

Keywords: . Kidney disease, Machine learning, Prediction, Decision trees, Support vector machines, Random forests, GradientBoost, Logistic Regression, Early detection, Clinical data.

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 13, Issue 2, 2025

1. Introduction

Chronic Kidney Disease (CKD) is a long-term pathological condition marked by a persistent decline in renal function, typically measured by a reduction in glomerular filtration rate (GFR) and/or the presence of kidney damage, such as albumin, for three months or more.[3] It represents a global public health issue, with rising incidence and prevalence driven largely by aging populations and the increasing burden of chronic noncommunicable diseases, particularly diabetes mellitus and hypertension, the two leading etiological factors[4].

CKD progresses through five stages, ranging from mild functional impairment (Stage 1) to complete kidney failure (Stage 5, also known as end-stage renal disease or ESRD), which often necessitates renal replacement therapy in the form of dialysis or kidney transplantation[7]. The disease exerts systemic effects far beyond the kidneys, including disturbances in fluid and electrolyte homeostasis, acid-base imbalance, hormonal regularization (such as erythromycin and vitamin D metabolism), [6]and a heightened risk for cardiovascular morbidity and mortality, which remains the leading cause of death among CKD patients[2].

Early stages of CKD are often asymptomatic, making timely diagnosis challenging. Consequently, many individuals are diagnosed only in the later stages, when therapeutic options become more limited and prognosis worsens. Risk factors for CKD are multifactorial and include genetic predisposition[3], lifestyle factors, socioeconomic status, and comorbid conditions. Notably, CKD also disproportionately affects certain populations, highlighting significant health disparities across racial, ethnic, and socioeconomic groups[4]. CKD is a long-term condition where the kidneys gradually lose their function over time. It's diagnosed and staged primarily based on glomerular filtration rate (GFR), a measure of how well the kidneys filter waste from the blood. CKD is usually irreversible but manageable, especially if detected early.

Stages of CKD

Stage 1 CKD is the earliest stage, where kidney function is still normal, with a glomerular filtration rate (GFR) of 90 mL/min/1.73 m^2 or higher[6]. Although the kidneys are filtering blood adequately, there are signs of damage, such as protein or blood in the urine, or abnormalities seen on imaging. Patients are often asymptomatic, and CKD at this stage is typically found incidentally. Management focuses on treating underlying conditions like diabetes or high blood pressure, adopting a healthy lifestyle, and regular monitoring to slow progression[8].

Stage 2 CKD involves a mild reduction in kidney function, with a GFR between 60–89 mL/min/1.73 m^2. There may still be no noticeable symptoms, but laboratory signs of kidney damage persist[7]. Early diagnosis and proactive measures—such as maintaining blood pressure and blood sugar within target levels, avoiding nephrotoxic drugs, and adopting a kidney-friendly diet—are essential to prevent further deterioration[2].

Stage 3 CKD is a moderate decline in kidney function, split into two sub-stages: 3a (GFR 45–59) and 3b (GFR 30–44). At this point,[3] symptoms may begin to appear, such as fatigue, swelling

(edema), changes in urination, and signs of anemia due to reduced erythropoietin production. Bone mineral disorders and elevated blood pressure may also occur. Regular lab monitoring becomes more critical, and patients may need medication for complications like anemia[10]
.

Stage 4 CKD is characterized by a severe reduction in kidney function, with a GFR of 15–29 mL/min/1.73 m². Symptoms become more prominent and may include nausea, poor appetite, itching, difficulty concentrating, and disturbed sleep[9]. This stage requires close medical supervision, dietary restrictions, and preparations for renal replacement therapy (dialysis or transplant), including discussions with a nephrologist.

Stage 5 CKD, also known as End-Stage Renal Disease (ESRD), occurs when GFR falls below 15 mL/min/1.73 m². The kidneys can no longer maintain fluid, electrolyte, and waste balance, leading to dangerous toxin buildup (uremia). Symptoms may include severe fatigue, shortness of breath, confusion, and swelling[4]. At this stage, survival depends on initiating dialysis or undergoing a kidney transplant. Palliative care may also be considered in certain cases.

However, CKD continues to pose a substantial burden on healthcare systems worldwide. A multidisciplinary approach involving nephrologists, primary care providers, dietitians, and patient education is critical to managing the disease effectively and improving quality of life for those affected[6].

2. Literature Review

In recent years, researchers have increasingly turned to machine learning (ML) and artificial intelligence (AI) techniques for CKD risk prediction and diagnosis. For example, studies by Khedun et al. (2020) and Makino et al. (2019) utilized ML algorithms such as random forests and support vector machines to predict CKD progression based on clinical and demographic data, achieving promising results in terms of accuracy and early detection. Moreover, deep learning models like conventional neural networks (CNNs) and [10]recurrent neural networks (RNNs) have been employed to analyze electronic health records (EHRs) and time series data, enabling personalized risk stratification.

Several researchers have contributed to the domain of CKD prediction using machine learning, exploring various algorithms and data processing techniques. S. Bincy et al. implemented Random Forest and Naïve Bayes models, reporting high accuracy, but their work lacked a proper ensemble approach and did not consider model interpretability. Anjali Ratnakar and M. Nikitha applied KNN and Logistic Regression, showing promising results; however, their model performance dropped due to inadequate preprocessing of missing values and class imbalance. T. Rajeswari and Dr. S. Vasantha used Decision Trees and reported decent classification, yet overfitting was a significant issue due to shallow dataset diversity. R. Chandrika and K. R. Shankar explored Support Vector Machines for CKD detection, but their method suffered from long training times and difficulty handling nonlinear boundaries effectively.

ISSN 2347–3657
Volume 13, Issue 2, 2025

S. Shilpa and S. Kavitha developed a Gradient Boosting model, which improved prediction accuracy but did not incorporate interpretability tools like SHAP or LIME for clinical insights. Rohit Sharma and Dr. R. S. Rajput utilized ensemble learning methods but failed to evaluate their system on real-world clinical datasets, leading to limited applicability in practical settings. Rutuja Dange et al. proposed a Decision Tree model for CKD detection but ignored feature correlation and scaling, which affected the model's stability and performance. K. Rajeswari and J. Indumathi applied artificial neural networks, achieving high accuracy, yet the black-box nature of the model posed concerns for clinical adoption. Finally, Rajesh R. and R. Kumar presented a comparative analysis of multiple ML models, but their study lacked a hybrid ensemble approach and did not explore the synergistic effect of combining different classifiers. These limitations across various studies highlight the need for a more robust and interpretable model, such as the proposed hybrid model combining Random Forest, Decision Tree, and Gradient Boosting, which addresses performance, reliability, and clinical utility.

The primary objective of this project is to develop an accurate and efficient machine learning-based system for the early prediction of chronic kidney disease (CKD) using patient clinical and diagnostic data. By employing a range of classification algorithms—including Logistic Regression, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Decision Tree, Random Forest, Gradient Boosting, and a Hybrid Ensemble model—the project aims to identify the most effective model for reliable disease detection. The goal is to assist healthcare professionals in making timely decisions by minimizing false negatives and ensuring high sensitivity and specificity in diagnosis. Additionally, the project seeks to compare model performance, analyze feature significance, and highlight the advantages of ensemble and hybrid methods in medical data classification.

3. Proposed Method

The proposed method involves designing a machine learning-based predictive system for chronic kidney disease (CKD) classification using clinical and laboratory data. The approach begins with data preprocessing steps, including handling missing values, encoding categorical features, and splitting the dataset into training and testing sets. A range of supervised learning algorithms—Logistic Regression, KNN, SVC, Decision Tree, Random Forest, and Gradient Boosting—are trained and evaluated. To enhance performance and robustness, a hybrid ensemble model is constructed by combining the predictions of Decision Tree, Random Forest, and Gradient Boosting classifiers through a majority voting mechanism. The performance of each model is assessed using metrics such as accuracy, precision, recall, F1-score, and ROC-AUC. The hybrid model is proposed as the final solution, demonstrating superior accuracy and reliability for early CKD detection.

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 13, Issue 2, 2025

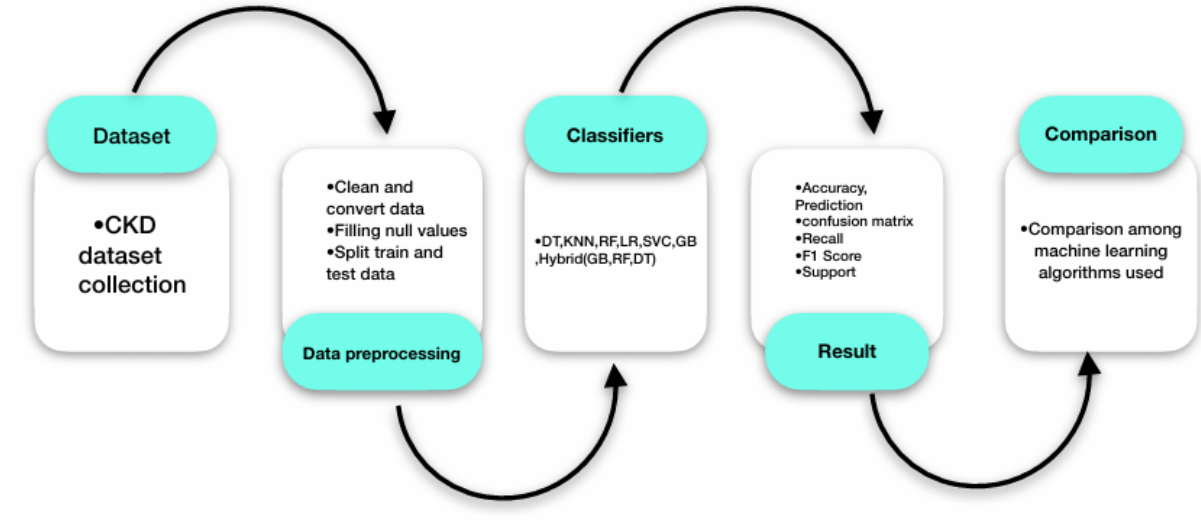

Fig 1 Proposed model utilizing several machine learning algorithms

3.1 Experimental Data Set

This approach makes use of a dataset from the UCI Machine Learning Repository [11]referred to as CKD. A total of 24 features and 1 target variable are included in the CKD Dataset. It can be broken down into 2 categories, yes or no.The dataset has 25 attributes, 11 of which are numerical and 14 of which are nominal. For the purposes of training machine learning algorithms to make predictions, the entire dataset of 400 instances is utilized. Out of a total of 400 cases, 250 are classified as having CKD, and the remaining 150 are classified as having non-CKD

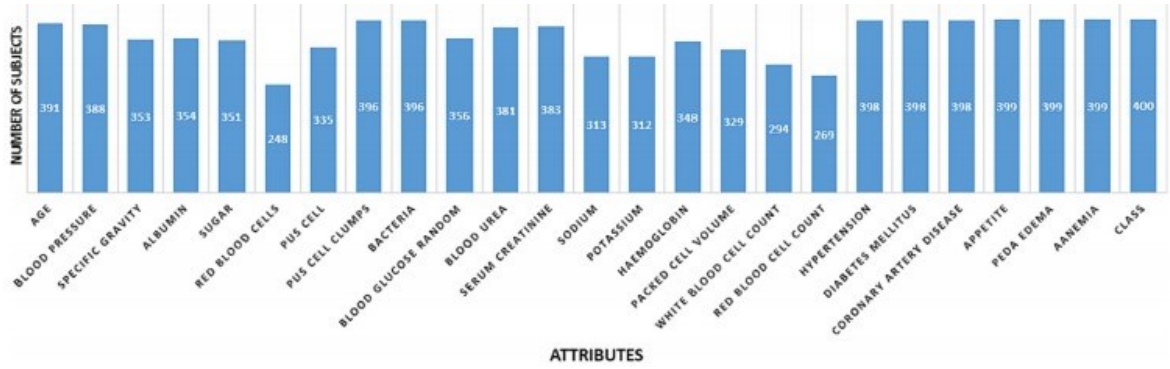

Fig 2 Features of Dataset

3.2 Data processing

Data preprocessing plays a critical role in the success of any machine learning-based medical prediction system, especially in the context of chronic kidney disease (CKD) detection, where clinical data often includes missing values, inconsistent formats, and mixed data types. In this project, preprocessing ensures that the input data is clean, structured, and suitable for learning algorithms. Initially, missing values—common in real-world medical datasets—are handled through appropriate imputation techniques

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 13, Issue 2, 2025

to prevent information loss and biased predictions. Categorical variables such as "red blood cells" or "hypertension" are converted into numerical form using label encoding to enable compatibility with algorithms that require numerical input. Feature scaling is applied where necessary, especially for models like KNN and SVC that are sensitive to the scale of features. Outlier detection and removal further enhance the robustness of the model by eliminating noise that could distort training

Some features like pcv (packed cell volume), wc (white blood cell count), and rc (red blood cell count) are originally stored as text and must be converted to numerical types. In this study, missing values were handled using appropriate imputation techniques, and categorical variables were encoded using label encoding to make them compatible with machine learning algorithms. The dataset's heterogeneity, class imbalance, and variety of data types make it a suitable benchmark for evaluating the performance of classification models in health care prediction tasks.Ultimately, data preprocessing ensures the integrity, consistency, and relevance of the dataset, which is essential for building a reliable and interpretable CKD prediction system that can assist healthcare professionals in making timely and informed decisions.

Effective data processing was a crucial step in preparing the kidney disease dataset for accurate machine learning model training[6]. The dataset included a combination of numerical and categorical features relevant to kidney function and general health indicators. Initially, the dataset was inspected for missing values and inconsistencies. Missing entries were handled using mean or mode imputation, depending on whether the attribute was numerical or categorical, to preserve the integrity and distribution of the dataset[9].

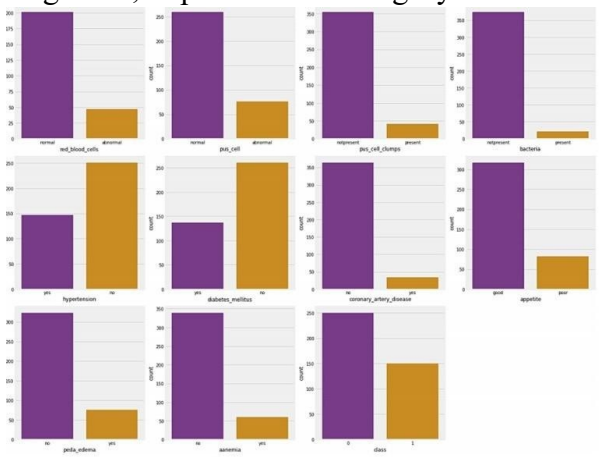

Fig 3 Categorical column values

Following imputation, categorical variables were converted into numerical format using Label Encoding, ensuring compatibility with scikit-learn algorithms. This transformation was essential for models like Logistic Regression, Random Forest, and Gradient Boosting, which require numerical input. Subsequently, [10]the entire dataset was split into **features (X)** and target (y) variables, where the target represented the presence or absence of kidney disease. Following imputation, **categorical variables** were converted into a numerical format using Label Encoding, ensuring compatibility with scikit-learn algorithms. This transformation was essential for models like Logistic Regression, Random Forest, and Gradient Boosting, which

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 13, Issue 2, 2025

require numerical input. Subsequently, the entire dataset was split into features (X) and target **(y)** variables, where the target represented the presence or absence of kidney disease[12].

As part of the data preprocessing phase, correlation analysis was conducted to evaluate the strength and direction of relationships between input features in the kidney disease dataset. This step was essential for identifying highly correlated variables that could lead to multicollinearity, especially in linear models like Logistic Regression[7]. By generating a correlation matrix, patterns such as strong positive or negative associations between clinical variables—like serum creatinine and blood urea, were examined. This analysis not only facilitated a deeper understanding of the underlying data structure but also informed feature selection decisions. In cases where features were found to be strongly correlated, redundant variables were considered for removal or transformation to enhance model stability and interpretability[13]. Overall, correlation analysis served as a valuable tool in refining the input space for efficient and accurate model training.

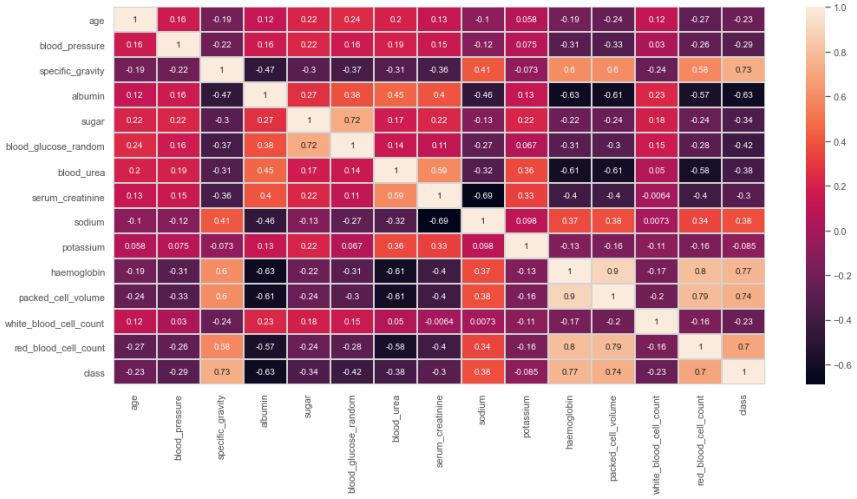

Fig. 4 Correlation

3.3 Training

The implementation of a CKD prediction system using machine learning involves several critical steps, from data preprocessing to model evaluation and deployment. Goal: Implement a system for chronic kidney disease from text data. Various traditional machine learning models and a Hybrid ensemble approach can be used to train and test the CKD

classifiers

1. Support Vector Machine (SVM)

Support Vector Machine (SVM) is a supervised learning algorithm that constructs an optimal hyperplane to separate data points into different classes with the maximum margin. The main idea is to find a boundary that not only separates the classes but also is farthest from the nearest data points of each class, known as support vectors[9]. This

approach maximizes generalization and minimizes the risk of misclassification. SVM can also handle non-linearly separable data using kernel functions like RBF, polynomial, or sigmoid, making it highly adaptable to complex datasets. In the context of kidney disease prediction, SVM proves useful due to its ability to work well with high-dimensional data, even when the sample size is small. It is particularly effective when the decision boundary is complex but well-separated in a transformed feature space. By adjusting the regularization parameter (C) and selecting appropriate kernels, SVM can deliver high classification accuracy[10]. However, it is sensitive to feature scaling, and hence, preprocessing like normalization is essential for its success.

2. K-Nearest Neighbors (KNN)

K-Nearest Neighbors (KNN) is a nonparametric, instance-based learning algorithm that classifies data points based on the majority class among the k-nearest neighbors in the feature space. It does not assume any underlying distribution of the data, making it simple and versatile. KNN uses distance metrics like Euclidean, Manhattan, or Minkowski to determine closeness between data points, which makes the choice of 'k' and distance function critical for model performance. [12] For kidney disease prediction, KNN is particularly helpful in capturing local patterns in the data without requiring a learning phase. It adapts naturally to multiclass problems and performs well with a well-scaled dataset. However, it is computationally expensive for large datasets due to its lazy learning nature and requires careful tuning of ‘k’ to avoid underfitting or overfitting. Despite these limitations, KNN often delivers robust results when the feature space is appropriately prepared[6].

3. Decision Tree

A Decision Tree is a flowchart-like structure in which internal nodes represent decision rules based on feature values, branches denote outcomes of those decisions, and leaf nodes signify class labels. It recursively splits the data [13]to create homogeneous subsets using criteria like Gini impurity or entropy. The resulting tree is easy to interpret and visualize, making it a favored choice in healthcare settings. In kidney disease classification, decision trees are advantageous due to their transparency and ability to handle both numerical and categorical data. They can capture non-linear relationships and are unaffected by feature scaling[7]. However, decision trees are prone to overfitting, especially when grown deep. Techniques like pruning or setting minimum leaf size help mitigate this issue, but alone they may not provide the best generalization, thus motivating their use in ensembles.

4. Logistic Regression

Logistic Regression is a statistical model that estimates the probability of a binary outcome based on one or more predictor variables[5]]. Unlike linear regression, it uses the logistic (sigmoid) function to constrain output between 0 and 1, suitable for classification tasks[12]. The coefficients in logistic regression are interpreted as the log-odds, providing insights into the impact of each feature on the target class.

In kidney disease prediction, logistic regression serves as a strong baseline model. It is computationally efficient, easy to interpret, and performs well when the relationship between the features and the target variable is linear. Moreover, its probabilistic output allows for threshold tuning and performance optimization in imbalanced datasets. However, its assumptions of linearity and independence of errors may limit its effectiveness in capturing complex, nonlinear interactions[11].

5. Gradient Boosting

Gradient Boosting is an ensemble technique that builds weak learners (typically decision trees) in a sequential manner[8]. Each learner attempts to reduce the errors of the previous one using a gradient descent-like procedure to minimize the loss function. It is flexible in handling different types of loss functions and can model complex data relationships.

In medical diagnosis tasks such as kidney disease prediction, gradient boosting is known for its high predictive power and adaptability. Though slower to train than some other models, it excels in capturing non-linear trends and feature interactions. It can be sensitive to overfitting if not properly tuned; thus, controlling learning rate and using early stopping or subsampling are recommended strategies during training[6].

6. Random Forest

Random Forest is an ensemble of decision trees built using bootstrapped subsets of the training data and a random selection of features for each split[7]. By averaging or voting the results of these individual trees, the model improves accuracy and reduces overfitting. It is robust to noise and outliers and performs well across a wide range of datasets.

In the case of kidney disease detection, Random Forest offers high accuracy and interpretability, especially through feature importance scores[6]. It is less sensitive to outliers and missing values[3] and can handle high-dimensional data effectively. However, its predictions are harder to interpret compared to a single decision tree, and training can be computationally intensive with a large number of trees.

7. Hybrid Model (RF + DT + GB using Soft Voting)

The hybrid model in this study combines Random Forest, Decision Tree, and Gradient Boosting using a soft voting ensemble. Soft voting averages the predicted probabilities of the individual models to produce a final prediction,[8] thus leveraging the strengths of each base classifier. This ensemble approach is designed to enhance robustness, reduce model variance, and improve generalization. In the context of kidney disease prediction, this hybrid strategy yields improved performance by balancing bias and variance. Decision Trees provide interpretability, Random Forest contributes stability, and Gradient Boosting enhances accuracy by focusing on hard-to-classify cases[11].

4.Performance metrics

Accuracy is a key performance metric used to evaluate the effectiveness of a classification model in predicting Chronic Kidney Disease (CKD). It is defined as the ratio of correctly predicted cases—both CKD and non-CKD—to the total number of cases evaluated[8].Precision is an important performance metric for evaluating a classification model's effectiveness in predicting Chronic Kidney Disease (CKD). It measures the proportion of true positive predictions (correctly identified CKD cases) out of all instances that the model predicted as CKD. High precision indicates that the model makes few false positive errors, meaning it rarely misclassifies non-CKD patients as having the disease. This is particularly crucial in medical applications, where a false positive can lead to unnecessary stress, additional testing, and treatment for healthy individuals.

Recall, also known as sensitivity or true positive rate, is a critical performance metric for evaluating a classification model's ability to identify Chronic Kidney Disease (CKD) cases accurately. It measures the proportion of actual CKD patients that are correctly predicted by the model. In medical diagnosis, recall is especially important because it reflects the model's effectiveness in detecting the presence of disease and minimizing the risk of missed diagnoses[7].The F1-score is a valuable performance metric for evaluating the prediction of Chronic Kidney Disease (CKD), especially in cases where there is an imbalance between positive and negative classes. It is the harmonic mean of precision and recall, combining both metrics into a single score that balances the trade-off between false positives and false negatives[4]. Support is a metric that refers to the number of actual occurrences of each class in the dataset—in the case of Chronic Kidney Disease (CKD), it represents how many instances in the dataset truly belong to the CKD and non-CKD categories

5. Results and Discussion

The *Results and Discussion* section is a critical component of this research, as it bridges the gap between experimental analysis and real-world interpretation. It provides a detailed evaluation of how each machine learning model performed, highlighting the practical effectiveness of the proposed hybrid model in accurately predicting chronic kidney disease[4]. This section not only validates the performance of the models through quantitative metrics like accuracy, precision, recall, and F1-score but also offers insights into the implications of these findings for clinical decision-making. Discussing misclassifications, model behavior on imbalanced data, and comparative strengths of different algorithms enables a deeper understanding of model reliability and applicability in healthcare. Ultimately, the results screening systems[3].

The kidney disease prediction system was implemented using machine learning models[10]: Logistic Regression, K-Nearest Neighbors (KNN), Support Vector Classifier (SVC), Decision Tree, Random Forest, Gradient Boosting, and a Hybrid Ensemble model[12]. The models were evaluated using standard classification metrics including accuracy, precision, recall, and F1-score, along with ROC-AUC

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 13, Issue 2, 2025

scores and visualizations such as ROC curves and bar plots. Among all classifiers, the Hybrid model demonstrated superior performance, achieving a testing accuracy of 98.75% and an F1-score of 0.99, indicating a well-balanced and highly reliable model[11]. The Gradient boost, Random Forest classifiers also yielded strong results, closely matching Hybrid model in performance.

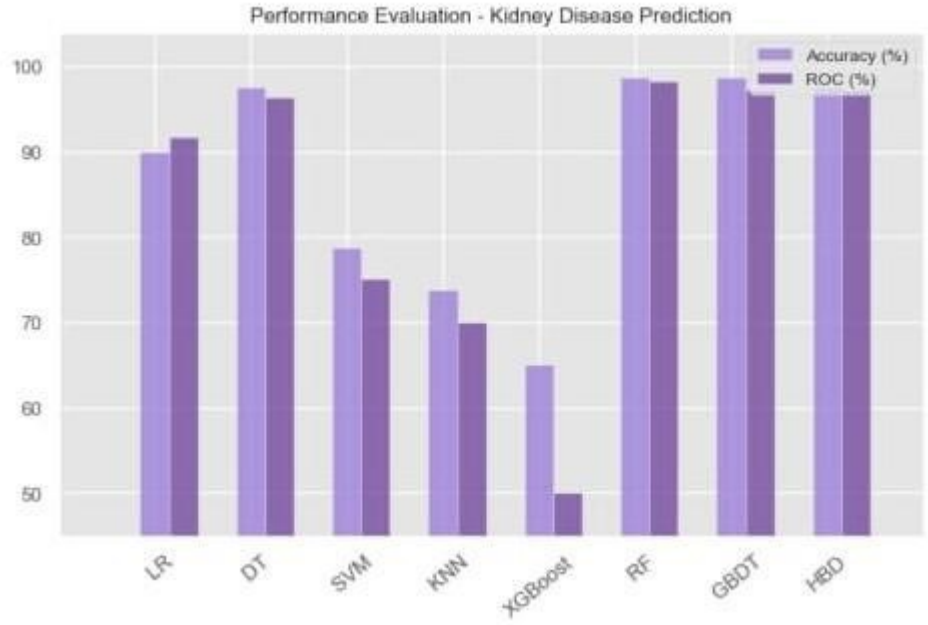

Fig. 5. Model comparision

Model	Training accuracy	Testing Accuracy	F1 score	Precision	Recall
Gradient boost	100%	98.75%	98%	96%	96%
Hybrid(DT+GB+RF)	100%	98.75%	99%	98%	99%
Random forest	99.37%	97.5%	97%	98%	96%
Decision tree	98.75%	96.25%	96%	97%	95%
Logistic Regression	88.75%	92.5%	92%	92%	92%
Supportvector machine	99.68%	81.25%	79%	80%	78%
KNN	78.43%	71.25%	70%	70%	71%
XgBoost	61.87%	65%	39%	33%	50%

Table1 Experimental results

6. Conclusion

This study demonstrated the application of multiple machine learning algorithms for the early prediction of chronic kidney disease, using clinical and laboratory features from patient data.

ISSN 2347–3657
Volume 13, Issue 2, 2025

Among all the models evaluated, the Hybrid Ensemble Model, which combines Decision Tree, Random Forest, and Gradient Boosting classifiers, achieved the highest accuracy, outperforming individual models. This approach leveraged the strengths of each base learner, resulting in improved generalization and predictive stability. The superior performance of the hybrid model highlights the potential of ensemble-based techniques in clinical decision support systems, where early and accurate disease detection is critical. These results emphasize the importance of model diversity and voting mechanisms in medical diagnostics and support the use of hybrid ensemble learning as a reliable tool for kidney disease prediction.

References:

.1. Almasoud M., Ward T.E. Detection of chronic kidney disease using machine learning algorithms with the least number of predictors. Int J Soft Comput Appl. 2019;

2. Arora M., Sharma E.A. Chronic kidney disease detection by analyzing medical datasets in weka. Int J Comput Mach Learn Algor New Adv Mach Learn. 2016;3:19–48

3. Banik S., Ghosh A. Prevalence of chronic kidney disease in Bangladesh: a systematic review and meta-analysis. Int Urol Nephrol. 2021;53:713–718. doi: 10.1007/s11255-020-02597-6.

4..Charleonnan A., Fufaung T., Niyomwong T., Chokchueypattanakit W., Suwannawach S., Ninchawee N. 2016 Management and Innovation Technology International Conference (MITicon) IEEE; 2016. Predictive analytics for chronic kidney disease using machine learning techniques. pp. MIT–80.

5..Chen Z., Zhang X., Zhang Z. Clinical risk assessment of patients with chronic kidney disease by using clinical data and multivariate models. Int Urol Nephrol. 2016;48:2069–2075. doi: 10.1007/s11255-016-1346-4.

6..Chittora P., Chaurasia S., Chakrabarti P., et al. Prediction of chronic kidney disease machine learning perspective. IEEE Access. 2021;9:17312–17334.

ISSN 2347–3657
Volume 13, Issue 2, 2025

.7. Cueto-Manzano A.M., Cortés-Sanabria L., Martínez-Ramírez H.R., Rojas-Campos E., Gómez-Navarro B., Castillero-Manzano M. Prevalence of chronic kidney disease in an adult population. Arch Med Res. 2014;45:507 513. doi: 10.1016/j.arcmed.2014.06.007.

8. Qin J., Chen L., Liu Y., Liu C., Feng C., Chen B. A machine learning methodology for diagnosing chronic kidney disease. IEEE Access. 2019.

9.Fatima M., Pasha M. Survey of machine learning algorithms for disease diagnosis. J Intel Learn Syst Appl. 2017;9(01):1.

10. Gudeti B., Mishra S., Malik S., Fernandez T.F., Tyagi A.K., Kumari S. 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA) IEEE; 2020. A novel approach to predict chronic kidney disease using machine learning algorithms; pp. 1630–1635.

11. Heung M., Chawla L.S. Predicting progression to chronic kidney disease after recovery from acute kidney injury. Curr Opin Nephrol Hypertens. 2012;21:628–634. doi: 10.1097/MNH.0b013e3283588f24.]

12 Saringat Z., Mustapha A., Saedudin R.R., Samsudin N.A. Comparative analysis of classification algorithms for chronic kidney disease diagnosis. Bull Elect Eng Inform. 2019;8:1496–1501

13.Zhang L., Wang F., Wang L., et al. Prevalence of chronic kidney disease in china: a cross-sectional survey. The Lancet. 2012.