Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol.46 No. 3 (2025)

Performance Evaluation of Advanced Machine Learning Models for Chronic Kidney Disease Classification and Prediction

D P Singh

Amity University Uttar Pradesh Greater Noida Campus

Abstract:

Chronic Kidney Disease (CKD) is a major global health concern, with increasing prevalence and significant impacts on patient morbidity and mortality. Early and accurate prediction and classification of CKD can greatly assist in timely interventions, improving patient outcomes and resource allocation in healthcare. This paper presents an extensive analysis of nineteen machine-learning models for prediction and classification in CKD. The performance of each model is assessed using confusion matrix metrics such as Accuracy, Precision, Recall, and F1-Score, offering a thorough evaluation of their predictive abilities. Moreover, ROC curves and AUC scores are employed to evaluate and compare the models' ability to distinguish between classes. Results indicate that certain models are more effective in handling the unique characteristics of CKD data, providing insights into how algorithm selection affects predictive performance. The Random Forest classifier outperformed the others, achieving an accuracy of 0.99, along with precision, recall, and F1-score values all at 0.99. This paper underscores the potential of machine learning advancements and predictive modeling in creating innovative solutions, particularly for improving prediction accuracy in kidney disease and similar fields. Also offers a comprehensive resource for selecting suitable machine learning models in CKD applications and highlights areas for future research in CKD-related predictive analytics.

Keywords: Chronic Kidney Disease, Machine Learning, Prediction, Classification, Healthcare Analytics, Accuracy, Precision, Recall, F1 Score.

1. Introduction: Chronic Kidney Disease (CKD) represents a major and escalating global public health issue, affecting millions of people and often advancing to end-stage renal disease and other serious health complications. As reported by the Global Burden of Disease Study, CKD ranked as the 18th leading cause of death worldwide in 2010, up from 27th in 1990 (Mills et al.)[5]. The condition is widespread and affects populations across all regions, irrespective of geographical location or socio-economic status (Jha et al.)[4]. Chronic Kidney Disease (CKD) develops because of various underlying health conditions, including diabetes, hypertension, and cardiovascular disorders (Webster et al.)[16]. Moreover, demographic characteristics such as age and gender also affect an individual's vulnerability to the disease. CKD patients may present with a diverse range of symptoms, such as abdominal and back pain, fever, vomiting, and skin rashes, among others. The disease stems from gradual and irreversible kidney damage, which compromises the organs' ability to filter blood and eliminate waste effectively. Classified as a “chronic” illness due to its slow progression, CKD has contributed to an 82% rise in years of life lost from kidney-related conditions, underscoring its increasing burden—now on par with that of diabetes (Levey et al.)[1].

Timely identification and precise classification of chronic kidney disease (CKD) are essential for enhancing patient care, allowing for early medical intervention and tailored treatment approaches (Zhang et al.)[37]. Conventional diagnostic techniques, such as measuring glomerular filtration rate (GFR) and conducting urinalysis, often lack effectiveness in detecting the disease during its initial stages (Levey & Coresh)[3]. As a result, healthcare systems worldwide are adapting by implementing advanced screening methods and generating reliable data on CKD prevalence, which is vital for informed healthcare planning and policy development (Couser et al.)[2].

In recent times, there has been a concerning increase in Chronic Kidney Disease (CKD) cases across both developed and developing countries. In developing regions, rapid urban growth has led to unhealthy lifestyle habits; contributing to a higher prevalence of diabetes and hypertension—key contributors to CKD (Jha et al.)[4]. For instance, in Pakistan, factors such as inadequate nutrition, polluted water, excessive use of medications, and self-medication practices are major drivers of the disease (Nasir et al.)[28]. On the other hand, in developed nations like the United States, over 26 million adults are affected by CKD. Here, risk evaluation is more structured and considers variables such as advanced age, female gender, anemia, cardiovascular conditions, and diabetes (CDC)[33].Chronic Kidney Disease (CKD) affects an estimated 5–15% of the global population and is responsible for around 5–10 million deaths each year (Luyckx et al.)[18]. Due to its progressive progression, timely and precise diagnosis is crucial for lowering death rates and managing long-term health issues. The situation is especially critical in resource-limited regions, where inadequate healthcare access, low awareness, and economic challenges significantly impede effective treatment (Stanifer et al.)[9].

The integration of machine learning (ML) techniques into healthcare has significantly enhanced the capabilities for early diagnosis and disease classification. ML algorithms are adept at handling large volumes of data, identifying complex patterns, and producing reliable predictive outputs, making them particularly valuable in the analysis of Chronic Kidney Disease (CKD) (Chandra et al.)[27]. Various studies have confirmed the efficiency of models like Support Vector Machines (SVM), Random Forests (RF), Artificial Neural Networks (ANN), and Decision Trees (DT) in detecting kidney-related conditions. Notably, SVMs using Gaussian radial basis function kernels and ANN architectures have achieved high accuracy and precision in CKD classification tasks (Raghu &Sriraam)[15]. Moreover, comparative research emphasizes the significance of employing dimensionality reduction strategies like Principal Component Analysis (PCA) and kernel tuning methods to optimize model performance (Selvaperumal&Sairam)[19].

Supervised classification methods are increasingly being adopted in medical diagnostics. Studies by researchers such as Wickramasinghe et al. [30] and Hsu et al. [12] highlight the effectiveness of classification algorithms like k-nearest neighbors (KNN), artificial neural networks (ANN), and support vector machines (SVM) in predicting chronic kidney disease (CKD). Notably, ANN models frequently outperform conventional statistical techniques. The growing emphasis on machine learning in healthcare is further reflected in the promising performance of models like backpropagation neural networks and Naive Bayes (NB) in disease prediction, as noted by Kavakiotis et al. [13].The critical need for accurate prediction tools is underscored by CKD's position among the leading causes of death worldwide. The World Kidney Day 2019 report estimates that kidney-related diseases result in over 2.4 million deaths each year, ranking CKD as the sixth most rapidly increasing cause of global mortality (ISN) [22]. In developing nations such as Ethiopia, challenges such as inadequate access to clean water, proper nutrition, and healthcare services intensify the impact, with kidney disease contributing to an increasing share of overall mortality (Tessema et al.) [36].

The National Kidney Foundation (NKF) classifies chronic kidney disease (CKD) into five stages, determined by declining kidney function and glomerular filtration rate (GFR). Stages 1 and 2 often show few noticeable symptoms, whereas stage 5 signifies complete kidney failure, necessitating Renal Replacement Therapy (RRT)—a costly and demanding treatment [29]. In settings with limited resources, CKD management is particularly challenging due to insufficient infrastructure and a shortage of skilled healthcare providers. Therefore, leveraging advanced technologies for early detection is crucial to alleviate healthcare strain and enhance patient outcomes.

This research employs Random Forest, Support Vector Machine, and Decision Tree algorithms to predict and classify Chronic Kidney Disease (CKD), extending beyond simple binary classification to assess disease severity a critical factor in determining treatment strategies. By harnessing machine learning, the study seeks to provide dependable alternatives to conventional diagnostic methods. This study conducts a comprehensive assessment of various machine learning models by analyzing essential performance metrics such as Accuracy, Precision, Recall, F1 Score, and Mean Cross-Validation Score. The objective is to determine the most efficient

models for diagnosing Chronic Kidney Disease (CKD), thereby providing valuable insights to assist healthcare professionals in choosing the best predictive tools for accurate and timely CKD management.

2. Literature Review: Chronic Kidney Disease (CKD) is a significant worldwide health issue, requiring timely and precise detection to avoid advancement to end-stage kidney failure. Machine Learning (ML) has become a valuable approach for enhancing CKD prediction and classification by analyzing patient health data. Various studies have explored different ML models and data pre-processing methods to improve diagnostic precision. In one study, Salekin and Stankovic [10] assessed several classifiers, such as K-Nearest Neighbors (KNN), Random Forest (RF), and Artificial Neural Networks (ANN), using a dataset of 400 patient records. Their method incorporated wrapper-based feature selection, reducing the dataset to five critical attributes. The Random Forest model outperformed others, reaching 98% accuracy with a Root Mean Square Error (RMSE) of 0.11.

Charleonnan et al. [11] evaluated various predictive models, including KNN, Support Vector Machine (SVM), Logistic Regression (LR), and Decision Tree (DT), using a dataset from Indian CKD patients. Their study sought to determine the best-performing classifier for predicting CKD. In another study, Tekale et al. [20] examined a CKD dataset containing 400 records and 14 carefully selected features post pre-processing. Their findings showed that SVM surpassed DT, achieving an accuracy rate of 96.75%. Additionally, Xiao et al. [26] investigated CKD progression by assessing multiple models such as LR, Elastic Net, Lasso, Ridge Regression, SVM, RF, XGBoost, Neural Network, and KNN. Their research utilized data from 551 proteinuria patients, categorizing the disease into mild, moderate, and severe stages. The results indicated that Logistic Regression performed best, with an AUC of 0.873, a sensitivity of 0.83, and a specificity of 0.82.

Priyanka and colleagues [23] investigated multiple algorithms such as Naïve Bayes, KNN, SVM, Decision Trees, and Artificial Neural Networks. Their findings showed that Naïve Bayes performed the best, reaching an accuracy of 94.6%. Similarly, Almasoud and Ward [21] employed Pearson correlation, ANOVA, and Cramer's V test to select features before training models with Logistic Regression, SVM, Random Forest, and Gradient Boosting (GB). Among these, GB demonstrated superior performance with an F-measure of 99.1%. In a separate study, Yashfi [31] utilized Random Forest and ANN to predict chronic kidney disease (CKD) risk based on 20 selected features, with Random Forest achieving the highest accuracy of 97.12%. Rady and Anwar [25] evaluated Probabilistic Neural Network (PNN), Multilayer Perceptron (MLP), SVM, and Radial Basis Function (RBF), finding that PNN delivered the best accuracy at 96.7%. Additionally, Alsuhibany et al. [32] introduced an Ensemble Deep Learning-based Clinical Decision Support System (EDL-CDSS) designed specifically for CKD diagnosis in IoT-enabled healthcare environments.

Poonia et al. [35] employed various machine learning algorithms—such as KNN, ANN, SVM, Naïve Bayes, and Logistic Regression—along with feature selection techniques like Recursive Feature Elimination (RFE) and the Chi-Square test. Their work, based on publicly accessible datasets, underscored the significance of effective pre-processing and model selection in enhancing CKD prediction. Similarly, Vinod [34] evaluated seven ML algorithms on a CKD dataset and identified KNN as the top performer, achieving 97% accuracy. Additional studies include those by Singh [39], who identified Logistic Regression as the most effective model for diabetes prediction, and Singh [40–43], who investigated ML approaches for forecasting breast cancer, lung cancer, and cardiovascular diseases. These works consistently show that ensemble methods such as AdaBoost, CatBoost, XGBoost, and Random Forest yield strong classification results, often surpassing 97% accuracy.

Various data mining techniques have been widely utilized for predicting Chronic Kidney Disease (CKD). Polat et al. [14] combined classifiers such as KNN, Naïve Bayes, SVM, preference tables, Random Forest (RF), and J48 with ensemble methods and k-means clustering to investigate treatment options and diagnostic processes. Ani et al. [6] employed Decision Trees (DT), Naïve Bayes, Linear Discriminant Analysis (LDA), and Backpropagation Networks to improve prediction accuracy. Wickramasinghe et al. [17] highlighted the importance of patient medical histories in developing a computational framework aimed at recommending suitable dietary plans for CKD patients. Arora and Sharma [7] proposed a feature selection approach using WEKA tools, while Eroğlu and Palabaş [8] examined indexing techniques to detect recurring kidney infections.

Qin et al. [24] enhanced model performance by applying KNN imputation to handle missing values in the UCI CKD dataset.

Research highlights diverse applications of machine learning in diagnosing chronic kidney disease (CKD), underscoring the importance of high-quality datasets, effective feature selection, appropriate algorithms, and model optimization. Predictive accuracy often depends on these factors. Although advancements have been achieved, further comparative studies are needed to comprehensively assess multiple ML models using varied datasets and evaluation metrics.

This study seeks to address this gap by performing a thorough evaluation of multiple machine learning algorithms for the early detection and classification of chronic kidney disease (CKD). Using a detailed dataset sourced from Kaggle (https://www.kaggle.com), we examine how clinical features correlate with CKD progression. Our methodology combines predictive modeling and advanced data analytics to create precise and effective classification models, ultimately enhancing clinical decision-making in CKD diagnosis.

3. Proposed methodology: Our proposed approach seeks to determine the most effective algorithm for kidney disease prediction by assessing the performance of nineteen different machine-learning models. [38,39,40,41,42,43,44].

3.1. Logistic Regression: Logistic Regression models the probability of a binary outcome using a logistic (sigmoid) function. For a given input feature vector $x = [x_1, x_2, x_3 \ldots\ldots x_n]$

Model: $P\left(x = \frac{1}{X}\right) = \sigma(W^T X + p)$,where $\sigma(z) = \frac{1}{1+e^{-z}}$ (sigmoid function), $W = [w_1, w_2, w_3 \ldots\ldots w_n]$ are the model weights, b is the bias term, $P\left(y = \frac{0}{x}\right) = 1 - P(y = \frac{1}{x})$.

Decision Rule: $\hat{y} = \begin{cases} 1 & If\ P(y = \frac{1}{x}) \geq 0.5 \\ 0 & Otherwise \end{cases}$

3.2. Decision Tree Classifier: Decision Tree recursively splits the data based on feature values to maximize a purity criterion (e.g., Gini index or Information Gain).

Splitting Criterion: For a node with N samples:

Gini Index: $G = 1 - \sum_{i=1}^{k} p_i^2$,where P_i is the proportion of samples belonging to class i.

Information Gain: $= H_{Parent} - \sum_i \frac{N_i H_i}{N}$, where $H = -\sum p_i \log_2 p_i\ (Entropy)$

Decision Rule: Traverse the tree according to feature splits until a leaf node is reached. The class label of the leaf node is the prediction.

3.3. Random Forest Classifier: Random Forest is an ensemble of Decision Trees trained on random subsets of data and features.

Model: For T trees, each tree outputs a prediction $\hat{y}_t$. The final prediction is determined by majority voting: $\hat{y} = mode(\{\hat{y}_1, \hat{y}_2, \hat{y}_3, \ldots\ldots \hat{y}_T\})$.

Feature Randomness: Each tree uses a random subset of features at each split to reduce correlation between trees.

3.4. Support Vector Machine (SVM): SVM finds a hyperplane that maximizes the margin between two classes in a feature space.

Model: For a feature vector x: $f(x) = w^T x + b$, Where: w is the weight vector, b is the bias.

Optimization Problem: $\min_{w,b} \frac{1}{2}\|w\|^2$ subject to $y_i(w^T x_i + b) \geq 1, \forall\ i$.

Decision Rule: $\hat{y} = \begin{cases} 1 & If f(x) \geq 0 \\ -1 & Otherwise \end{cases}$

Kernel Trick: To handle non-linear data, SVM uses kernels $K(X_i, X_j)$ to map data into a higher-dimensional space.

3.5. K-Nearest Neighbors (KNN): KNN predicts the label of a sample based on the labels of its k nearest neighbors in the feature space.

Model: Given a distance metric $d(x_i,x_j)$, identify the k nearest neighbors of a query point x_q.

Decision Rule: $\hat{y} = mode(\{\hat{y}_1, \hat{y}_2, \hat{y}_3, \ldots\ldots \hat{y}_k\})$. Where are the labels of the k nearest neighbors.

Euclidean: $\sqrt{\sum_n (x_{i,n} - x_{j,n})^2}$

3.6. Naive Bayes (Gaussian): Model Assumptions: Features are conditionally independent given the class. Each feature follows a Gaussian (normal) distribution.

Formula: For a feature vector $X = [x_1, x_2, x_3 \ldots\ldots x_n]$ the posterior probability is:

$P(C_k|X) = \frac{P(C_k)\prod_{i=1}^{n} P(x_i|C_k)}{P(X)}$, where $P(x_i|C_k)$is modelled as: $P(x_i|C_k) = \frac{1}{\sqrt{2\pi\sigma_{k,i}^2}} \exp\left(-\frac{(x_i-\mu_{k,i})^2}{2\pi\sigma_{k,i}^2}\right)$

With $\mu_{k,i}$ and $\sigma_{k,i}^2$ being the mean and variance of feature i for class k.

3.7. Naive Bayes (Bernoulli): Model Assumptions: Features are binary.Features are conditionally independent given the class.

Formula: For binary features X: $P(C_k|X) = \frac{P(C_k)\prod_{i=1}^{n} P(x_i|C_k)^{x_i}(1-P(x_i|C_k))^{1-x_i}}{P(X)}$

3.8. XGBoost (Extreme Gradient Boosting): Model: XGBoost is an ensemble method based on decision trees optimized with gradient boosting.

Objective Function: $\mathcal{L} = \sum_{i=1}^{n} l(y_i, \hat{y_i}) + \sum_{k=1}^{K} \Omega(f_k)$ where: $l(y_i, \hat{y_i})$is the loss function (e.g., log loss for classification).

$\Omega(f_k) = \mathrm{Y}T + \frac{1}{2}\lambda\|w\|^2$a regularization term to penalize tree complexity.

Prediction: $\hat{y_i} = \sum_{k=1}^{K} f_k(X_i)$ where f_k is the k-th decision tree.

3.9. Cat Boost: Model: CatBoost is a gradient-boosting algorithm specifically designed for categorical features.

Objective Function: Same as XGBoost, but with additional handling of categorical features using "ordered boosting."

Handling Categorical Features: Transform categorical features into numeric representations using statistics derived from the training data.

3.10. AdaBoost Classifier: Model: AdaBoost combines weak learners (typically decision stumps) iteratively.

Prediction: $H(X) = sign(\sum_{t=1}^{T} \alpha_t h_t(X))$, where:$h_t(X)$: Weak classifier at iteration t.α_t : Weight of h_t, determined by its accuracy.

Update Rule: Weights for misclassified samples are increased: $w_i, t+1 = w_i, t \exp(\alpha_t . 1_{y_i \neq h_t(x_i)})$

3.11. Extra Trees Classifier (Extremely Randomized Trees): Model: An ensemble of randomized decision trees.

Split Criteria: Unlike standard decision trees, Extra Trees randomly selects Features for split. Split thresholds within the feature range.

Prediction: $H(X) = \frac{1}{T}\sum_{t=1}^{T} h_t(X)$, where T is the number of trees.

3.12.Linear Discriminant Analysis (LDA): Objective: LDA is a linear classifier that assumes multivariate normal distribution of features within each class and equal covariance matrices across classes.

Modeling Steps: Prior Probability P(Y=c)): The probability of a given class c in the dataset.

Likelihood P(X|Y=c)): Features X follow a multivariate normal distribution:

where d is the dimensionality, μ_c is the mean vector of class c, and Σ is the shared covariance matrix.

Decision Rule (Discriminant Function): $\delta_c(X) = X^T\sum^{-1}\mu_c - \frac{1}{2}\mu_c^T\sum^{-1}\mu_c + \ln P\,(Y = c)$

Assign X to class c that maximizes $\delta_c(X)$.

3.13. Multi-Layer Perceptron (MLP): Objective: MLP is a specific type of feedforward neural network composed of multiple layers, including one or more hidden layers.

Mathematical Formulation: Layer-Wise Computation: For layer $:z^{(l)} = \sigma\left(W^{(l)}\alpha^{(l-1)} + b^{(l)}\right)$

where: $z^{(l)}$ is the pre-activation output,$a^{(l-1)}$ is the output of the previous layer, $W^{(l)}$ and $b^{(l)}$ are weights and biases for layer σ(·) is the activation function.

3.14. Stochastic Gradient Descent (SGD) Classifier: SGD minimizes a loss function L(w) using iterative updates with a learning rate η:Objective: $L(w) = \frac{1}{n}\sum_{i=1}^{n} l\,(y_i, f(x_i; w))$

Where: $l(y_i, f(x_i; w))$: The loss for the i-th sample. $f(x_i; w)$: The predicted output for x_i based on weights w.

Update Rule: $w^{t+1} = w^t - \eta\nabla l\left(y_i, f(x_i; w)\right)$, where: η: Learning rate. $\nabla l\left(y_i, f(x_i; w)\right)$ Gradient of the loss with respect to weights w, computed for a single sample i.

Common loss functions: Hinge loss for SVM: $l\left(y, f(x; w)\right) = \max(0,1 - w.f(x, w))$

Logistic loss for logistic regression: $l\left(y, f(x; w)\right) = \log(1 + \exp(-y.f(x, w)))$.

3.15. Bagging Classifier: Bagging (Bootstrap Aggregating) is an ensemble technique that combines the predictions of multiple base estimators (e.g., decision trees).

Objective: The final prediction f(x) is the aggregate of individual predictions from B base models: $f(x) = \frac{1}{B}\sum_{b=1}^{B} f_b\,(x)$. Where: $f_b(x)$ The prediction of the b-th base model. B: Total number of base models.

Workflow: Randomly sample m datasets with replacement from the original dataset. Train a base model $f_b(x)$ on each sampled dataset. Aggregate predictions by: Majority vote for classification. Averaging for regression.

3.16. Histogram-Based Gradient Boosting: This is a variant of Gradient Boosting that uses histograms to speed up the computation of splits.

Objective: Gradient Boosting builds an additive model by minimizing a loss function $L\left(y, f(x)\right): f(x) = \sum_{t=1}^{T}\alpha_t\, h_t(x)$, where: $h_t(x)$: The t-th weak learner. α_t: Weight of the t-th learner. T:Total number of learners.

Gradient Descent: Update at each stage: $h_t(x) = \arg\min_h \sum_{i=1}^{n}\left[\nabla\, \mathrm{L}\left(y_i, f^{t-1}(x_i)\right).h(x_i) + \frac{1}{2}\, h(x_i)^2\right]$

Where: $\nabla\, \mathrm{L}\left(y_i, f^{t-1}(x_i)\right)$: Gradient of the loss for i-th sample. Histograms are used to bucketize features, reducing computational complexity for split selection.

3.17. Ridge Classifier: Ridge Classifier is a linear model that applies L2 regularization to reduce overfitting.

Objective: Ridge regression minimizes the following loss: $L(w) = \frac{1}{n}\sum_{i=1}^{n}(y_i - w^T x_i)^2 + \lambda\|w\|_2^2$

Where: : True label for the i-th sample. : Feature vector for the i-th sample.w: Weight vector. λ: Regularization parameter.

The optimal weights w are found as: $w = (X^T X + \lambda I)^{-1} X^T y$

Where: X: Design matrix of features. y: Vector of target values.

For classification, the predicted class is determined by the sign of $w^T x$.

3.18. Gradient Boosting Machine (GBM): GBM builds an ensemble of decision trees sequentially to minimize the loss function of the model.

Mathematical Formulation: Objective: Minimize the loss L over the training data: $L(y, \hat{y}) = \sum_{i=1}^{n} l\,(y, f(x_i))$

where: y: True labels. $\hat{y}$: Predicted values. $f(x_i)$: Model output for input x_i

ℓ(y,f(x)): Chosen loss function (e.g., Mean Squared Error for regression or Log Loss for classification).

Boosting Process: The model is constructed iteratively as: $f_m(x) = f_{m-1}(x) + \vartheta h_m(x)$

where: $f_m(x)$: Ensemble model at iteration. $h_m(x)$: Weak learner (usually a decision tree) fitted to the negative gradient of the loss function. : Learning rate, a shrinkage parameter to control the contribution of each tree.

Gradient Step: At each iteration m, the weak learner minimizes the pseudo-residuals:

$$r_i^{(m)} = -\frac{\partial l(y_i, f(x_i))}{\partial f(x_i)}$$

The new weak learner $h_m(x)$ is trained to predict $r_i^{(m)}$. Final Prediction: After M iterations: $\hat{y} = f_M(x)$.

3.19. Stochastic Gradient Boosting Machine (SGBM): SGBM introduces randomness by training on a random subsample of the data at each iteration, reducing variance and preventing overfitting.

Modifications to GBM: 1.Subsampling: Instead of using the entire dataset to fit$h_m(x)$ a random subset $S \sqsubseteq \{1,2,\ldots.n\}$ is sampled without replacement.Subsample size is controlled by the parameter fraction ∈(0,1]. The weak learner is trained on this subset: $h_m(x)$ is trained on S.

2. Objective and Prediction: The remaining process (gradient calculation, model updating, etc.) is the same as in GBM.

Final model: $f_m(x) = f_{m-1}(x) + \vartheta h_m(x)$, S ~ Uniform Sampling.

4. Confusion Matrix in Machine Learning: A confusion matrix summarizes a machine learning model's performance on a test dataset, visually displaying both accurate and inaccurate predictions. It is commonly used to evaluate classification models, which assign categorical labels to input data. This matrix is essential for assessing a classification model's performance, providing detailed counts of true positives, true negatives, false positives, and false negatives. It enables a deeper understanding of the model's recall, accuracy, precision, and overall ability to distinguish between classes by showing the frequency of predicted outcomes on the test dataset[39,40,41,42,43,44].

4.1 Accuracy: Accuracy measures a model's effectiveness by calculating the ratio of correctly classified instances to the total number of instances. $Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$,

where TP= True positives, TN= True negatives, FP= False positives and FN= False negatives.

4.2 Precision: Precision refers to the accuracy of a model's positive predictions. It is measured by the ratio of true positive predictions to the total number of positive predictions made by the model. $Precision = \frac{TP}{TP+FP}$

4.3 Recall: Recall measures how well a classification model can identify all the relevant instances within a dataset. It is calculated by dividing the number of true positive (TP) cases by the total number of true positives and false negatives (FN). $Recall = \frac{TP}{TP+FN}$

4.4 Specificity: Specificity, an essential metric for evaluating classification models, particularly in binary cases, measures how accurately a model identifies negative instances, also known as the True Negative Rate. $Specificity = \frac{TN}{TP+FP}$

5. Data Pre-processing and Exploratory Analysis for Chronic Kidney Disease Prediction:

5.1 Data Source and Features: To evaluate chronic kidney disease, we utilized a dataset obtained from [https://www.kaggle.com/datasets/mansoordaku/ckdisease], comprising 24 clinical attributes. These attributes were employed to develop predictive models aimed at identifying heart disease. As shown in Table 1, certain features in the dataset contain missing values.

Table1:Detail of Data Frame

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 400 entries, 0 to 399
Data columns (total 25 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   age                     391 non-null    float64
 1   blood_pressure          388 non-null    float64
 2   specific_gravity        353 non-null    float64
 3   albumin                 354 non-null    float64
 4   sugar                   351 non-null    float64
 5   red_blood_cells         248 non-null    object
 6   pus_cell                335 non-null    object
 7   pus_cell_clumps         396 non-null    object
 8   bacteria                396 non-null    object
 9   blood_glucose_random    356 non-null    float64
 10  blood_urea              381 non-null    float64
 11  serum_creatinine        383 non-null    float64
 12  sodium                  313 non-null    float64
 13  potassium               312 non-null    float64
 14  hemoglobin              348 non-null    float64
 15  packed_cell_volume      330 non-null    object
 16  white_blood_cell_count  295 non-null    object
 17  red_blood_cell_count    270 non-null    object
 18  hypertension            398 non-null    object
 19  diabetes_mellitus       398 non-null    object
 20  coronary_artery_disease 398 non-null    object
 21  appetite                399 non-null    object
 22  pedal_edema             399 non-null    object
 23  anemia                  399 non-null    object
 24  class                   400 non-null    object
dtypes: float64(11), object(14)
memory usage: 78.2+ KB
```

5.2 Missing Values in Numerical Variables: A detailed analysis of missing data among numerical variables highlights several significant gaps that may affect the accuracy of predictive models. As shown in Table 2, red_blood_cell_count (131) and white_blood_cell_count (106) exhibit the highest number of missing values, followed by potassium (88), sodium (87), and packed_cell_volume (71). Variables like hemoglobin, sugar, specific_gravity, and albumin have a moderate level of missingness. In contrast, important clinical markers such as serum_creatinine, blood_urea, and age have comparatively fewer missing entries.

Table2:Missing Values for numerical columns

```
red_blood_cell_count      131
white_blood_cell_count    106
potassium                  88
sodium                     87
packed_cell_volume         71
hemoglobin                 52
sugar                      49
specific_gravity           47
albumin                    46
blood_glucose_random       44
blood_urea                 19
serum_creatinine           17
blood_pressure             12
age                         9
dtype: int64
```

Figure 1 illustrates the distribution of missing values among the numerical features in the dataset. The highest proportions of missing data are found in red_blood_cell_count (approximately 32%) and white_blood_cell_count (around 28%). Moderate missingness is observed in features like potassium, sodium, packed_cell_volume, and hemoglobin. In contrast, attributes such as blood_pressure, serum_creatinine, blood_urea, and age exhibit relatively low missing rates, each under 10%.

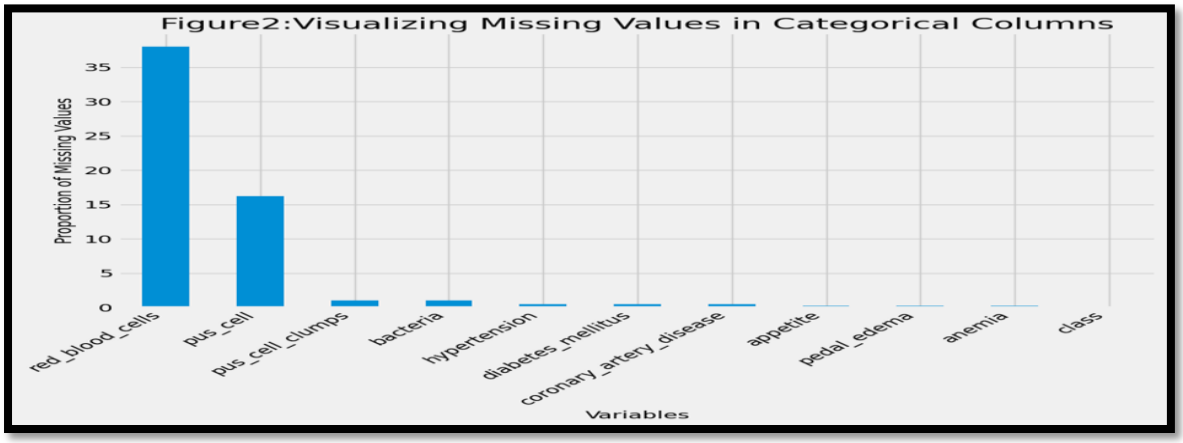
5.3 Missing Values in Categorical Variables: For the categorical variables (Table 3), the highest number of missing values are observed in red_blood_cells (152) and pus_cell (65). In contrast, other clinical features like hypertension, diabetes_mellitus, and anemia have only minor missing data. Notably, the target variable class has no missing values, maintaining the reliability of supervised learning tasks.

Table3: Missing values for categorical columns

red_blood_cells	152
pus_cell	65
pus_cell_clumps	4
bacteria	4
hypertension	2
diabetes_mellitus	2
coronary_artery_disease	2
appetite	1
pedal_edema	1
anemia	1
class	0

dtype: int64

Figure 2 illustrates the extent of missing data among categorical variables. Notably, red_blood_cells and pus_cell exhibit substantial missingness at approximately 36% and 16%, respectively, whereas other variables such as hypertension, diabetes_mellitus, and anemia are mostly complete.



5.4 Imputation Strategies for Numerical Features and Categorical Features: Considering the biological importance of many affected variables, it is crucial to apply suitable imputation methods—either statistical or model-driven—to retain the integrity of the data and ensure the reliability of the model. Techniques like mean, mode, KNN, or MICE are especially recommended for handling features with significant amounts of missing data.During the data pre-processing stage, missing numerical values were imputed using the mean. To facilitate computational processing, categorical (nominal) variables were converted into numerical codes. Specifically, the values 'normal' and 'abnormal' for the variables rbc and pc were encoded as 1 and 0, respectively. For pcc and ba, 'present' was coded as 1, and 'not present' as 0. Similarly, the variables htn, dm, cad, pe, and ane were encoded with 1 for 'yes' and 0 for 'no'. Lastly, the appetite variable was encoded as 1 for 'good' and 0 for 'poor'

5.5. Data Distribution and Class Imbalance: Figure 3 indicates that the majority of categorical variables have uneven class distributions, with certain categories significantly outnumbering others. The target variable also displays a class imbalance. To build accurate and fair classification models, it will be essential to address these disparities using resampling techniques or weighted learning approaches.

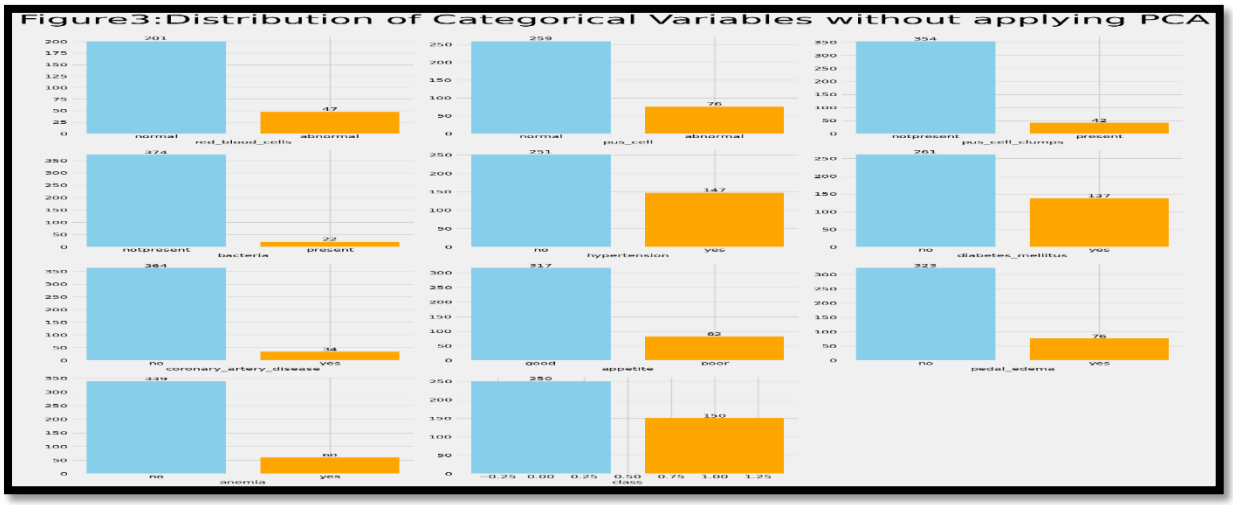
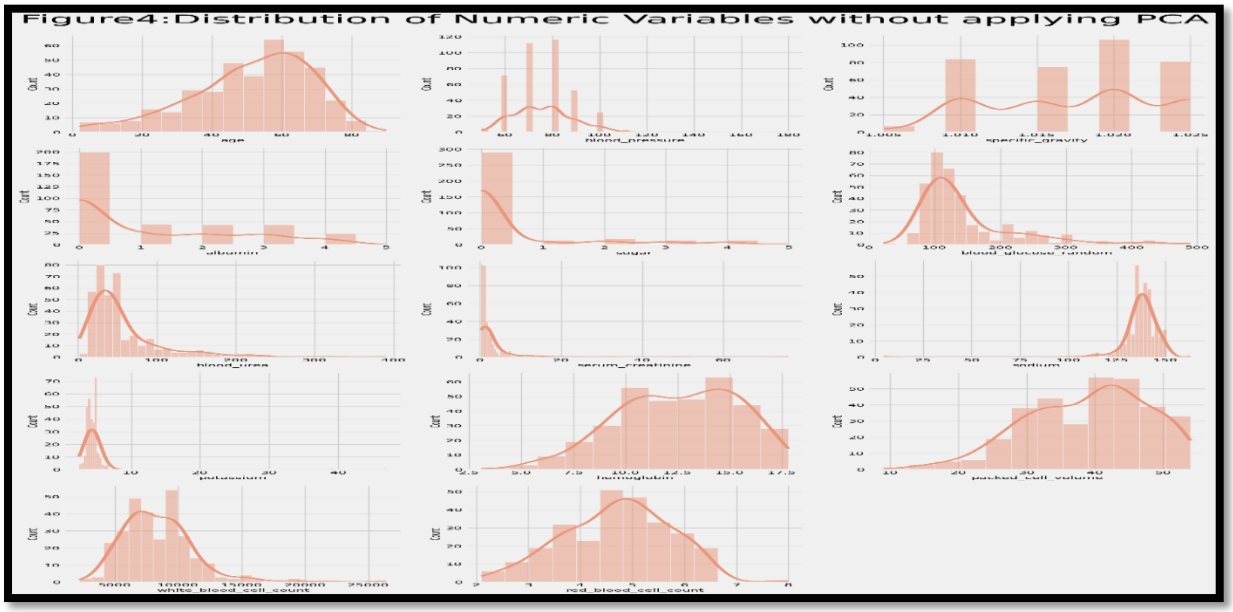
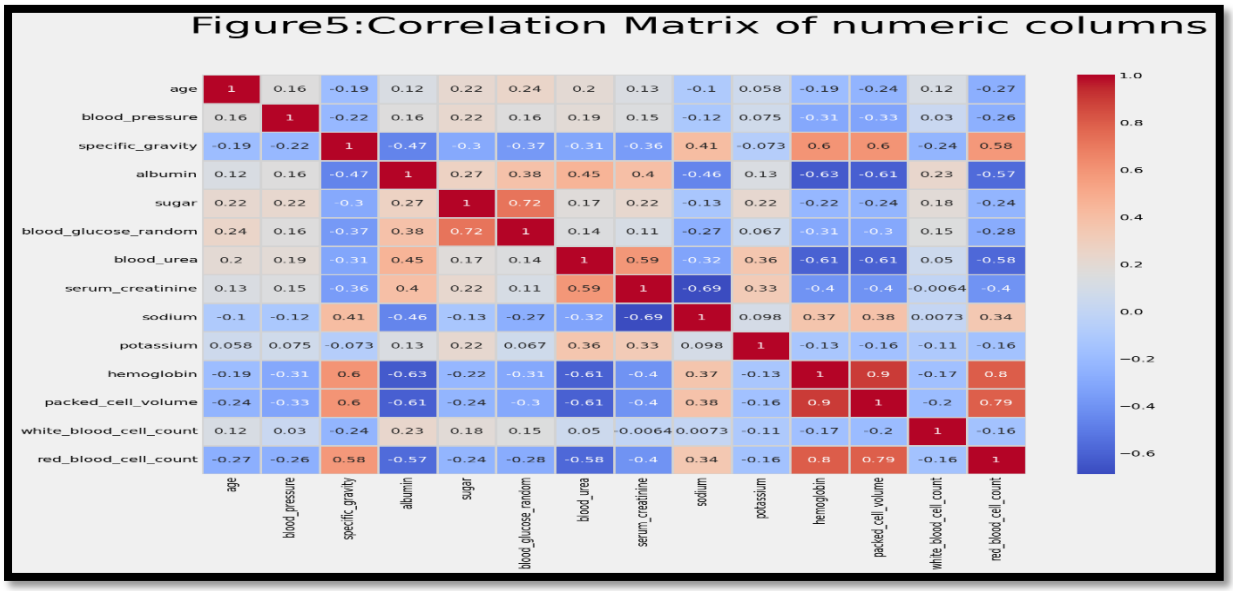


Figure 4 presents histograms and KDE plots to visualize the distribution of numeric variables. Certain features—such as blood_urea, serum_creatinine, potassium, sodium, and blood_glucose_random—display right-skewed distributions. Others, including specific_gravity, red_blood_cell_count, and white_blood_cell_count, reveal multimodal or irregular patterns. Variables like hemoglobin, packed_cell_volume, and age tend to follow a roughly normal distribution. Additionally, features such as sugar, albumin, and specific_gravity exhibit discrete or binned patterns, reflecting their ordinal nature. These distribution patterns highlight the importance of applying data transformation and scaling techniques to improve model effectiveness.



5.6 Correlation Analysis: Figure 5 presents the Pearson correlation matrix for the numerical variables, highlighting several notable relationships. Hemoglobin, packed cell volume, and red blood cell count exhibit strong positive correlations, suggesting both physiological association and potential redundancy. In contrast, strong negative correlations are observed between kidney function indicators (serum creatinine and blood urea) and nutritional markers (albumin and hemoglobin), underscoring their clinical significance. The majority of the remaining variables display weak to moderate correlations, indicating minimal multicollinearity.



6. Evaluation of Machine Learning Models for Chronic Kidney Disease Prediction: We assessed the effectiveness of nineteen machine-learning algorithms to evaluate their potential as clinical decision support tools for predicting chronic kidney disease.

6.1. Model Performance Comparison: To assess predictive performance, nineteen machine learning models were analyzed, and the most accurate model was identified. Among them, the Random Forest algorithm demonstrated the highest performance in predicting chronic kidney disease. It achieved an accuracy of 99.5%, precision of 99.8%, recall of 99.3%, and an F1 score of 99.4%. The detailed accuracy, precision, recall, and F1 scores for all models are provided in Table 4.

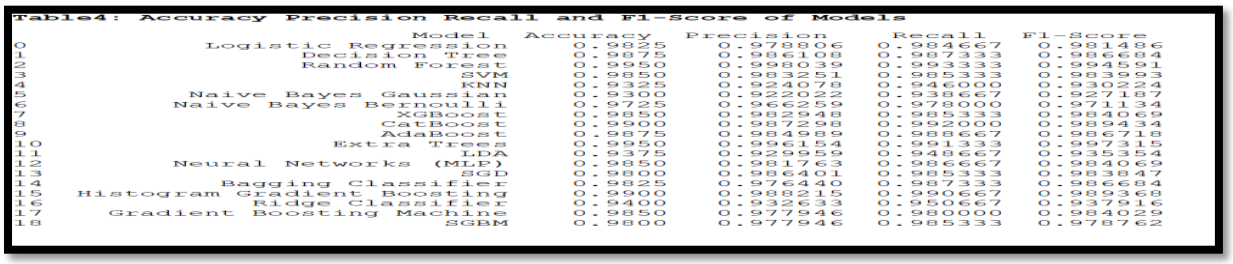
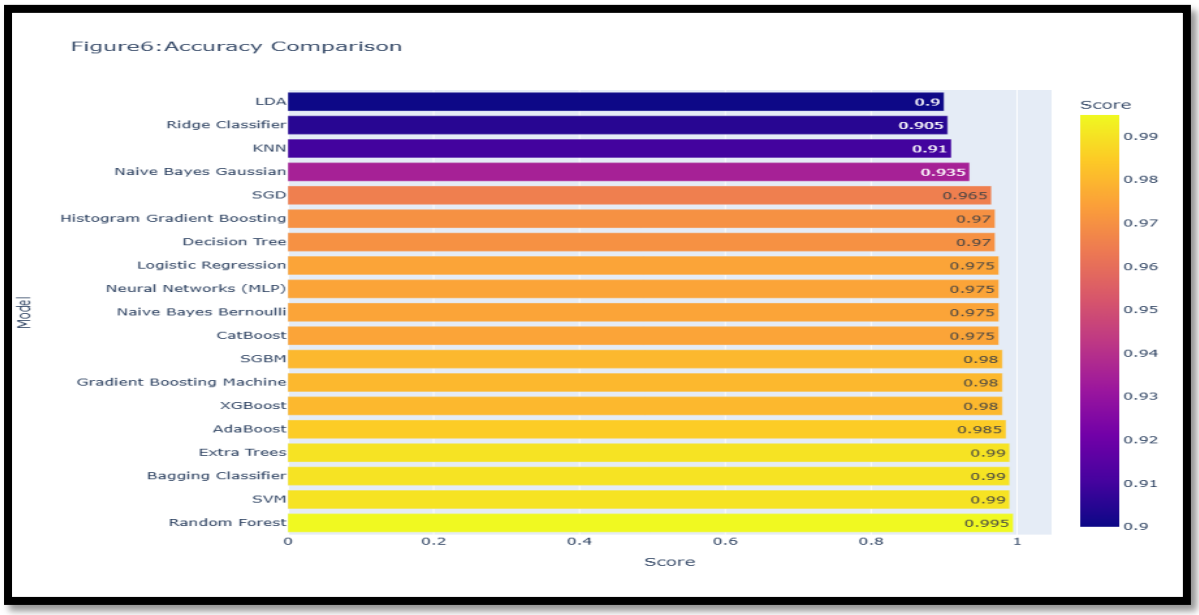


Table4: Accuracy Precision Recall and F1-Score of Models

	Model	Accuracy	Precision	Recall	F1-Score
0	Logistic Regression	0.9825	0.978806	0.984667	0.981486
1	Decision Tree	0.9875	0.986108	0.987333	0.986684
2	Random Forest	0.9950	0.998039	0.993333	0.994591
3	SVM	0.9850	0.983251	0.985333	0.983593
4	KNN	0.9325	0.924078	0.946000	0.935024
5	Naive Bayes Gaussian	0.9300	0.922022	0.938667	0.927187
6	Naive Bayes Bernoulli	0.9725	0.966259	0.978000	0.971134
7	XGBoost	0.9850	0.982948	0.985333	0.984069
8	CatBoost	0.9900	0.987298	0.992000	0.989434
9	AdaBoost	0.9875	0.984989	0.988667	0.986718
10	Extra Trees	0.9950	0.996154	0.994333	0.995735
11	LDA	0.9375	0.929959	0.948667	0.939354
12	Neural Networks (MLP)	0.9850	0.981763	0.986667	0.984069
13	SGD	0.9800	0.986401	0.985333	0.985847
14	Bagging Classifier	0.9825	0.976440	0.987333	0.981884
15	Histogram Gradient Boosting	0.9900	0.988215	0.990667	0.989368
16	Ridge Classifier	0.9400	0.932633	0.950667	0.937916
17	Gradient Boosting Machine	0.9850	0.977946	0.980000	0.984029
18	SGBM	0.9800	0.977946	0.985333	0.978762

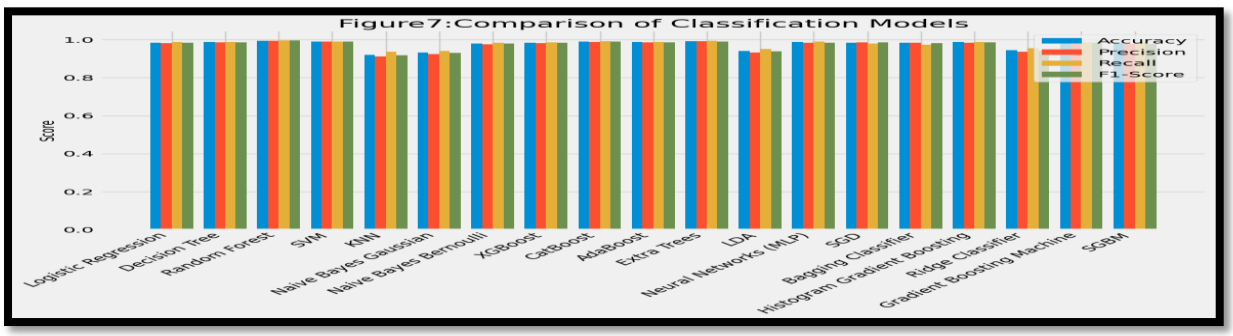
6.2 Accuracy Comparison Across Models: Figure 6 displays a horizontal bar chart comparing the accuracy scores of various classification models. Among them, the Random Forest algorithm achieved the highest performance with an impressive accuracy of 0.995. SVM, Bagging Classifier, and Extra Trees also showed remarkable results, each attaining an accuracy of 0.99. AdaBoost followed closely with a solid accuracy of 0.985, while CatBoost, XGBoost, and Gradient Boosting (GBM) all performed well, each recording an accuracy of 0.98.

Neural Networks (MLP), Bernoulli Naive Bayes, and Logistic Regression were moderate performers, each achieving a score of 0.975. Decision Tree and Histogram Gradient Boosting, both scoring 0.97, closely followed them. Lower performance was observed with the SGD Classifier (0.965) and Gaussian Naive Bayes (0.935), while KNN (0.91), Ridge Classifier (0.905), and LDA (0.90) recorded the weakest results.



6.3 Key Findings and Model Recommendations: The results suggest that ensemble methods such as Random Forest, Extra Trees, and Bagging, along with Support Vector Machines (SVM), offer the highest reliability for this task. In contrast, traditional approaches like LDA, Ridge, and KNN show relatively lower accuracy and might benefit from additional optimization or alternative methodologies. Although boosting techniques like AdaBoost, XGBoost, and CatBoost perform well, they slightly lag behind the leading models. In summary, priority should be given to tree-based ensemble methods and SVM for the best results, while less effective models may need further tuning or reconsideration.

6.4 Performance Metrics Analysis: Figure 7 presents a grouped bar chart illustrating the comparison of several classification models across four key performance indicators: Accuracy, Precision, Recall, and F1-Score. The Random Forest, Extra Trees, and Bagging Classifier models demonstrate superior accuracy and well-balanced results across all metrics, highlighting their robustness and dependability. Likewise, SVM and XGBoost show consistently strong performance, making them reliable alternatives. On the other hand, models such as the Naive Bayes variants and KNN exhibit noticeable discrepancies between precision and recall, suggesting a possible vulnerability to class imbalance. These models may require parameter optimization or dataset adjustments to improve their performance.The analysis highlights Random Forest, Extra Trees, Bagging Classifier, SVM, and XGBoost as the top-performing classifiers, consistently delivering strong and balanced results across all evaluation metrics. In comparison, other models may need further optimization, particularly when addressing challenges like class imbalance or complex feature relationships within the dataset.



6.5 Comprehensive Model Evaluation Using Multiple Metrics: Table 5 provides a detailed comparison of 19 classification models assessed across five crucial performance metrics: Accuracy, Precision, Recall, F1-Score,

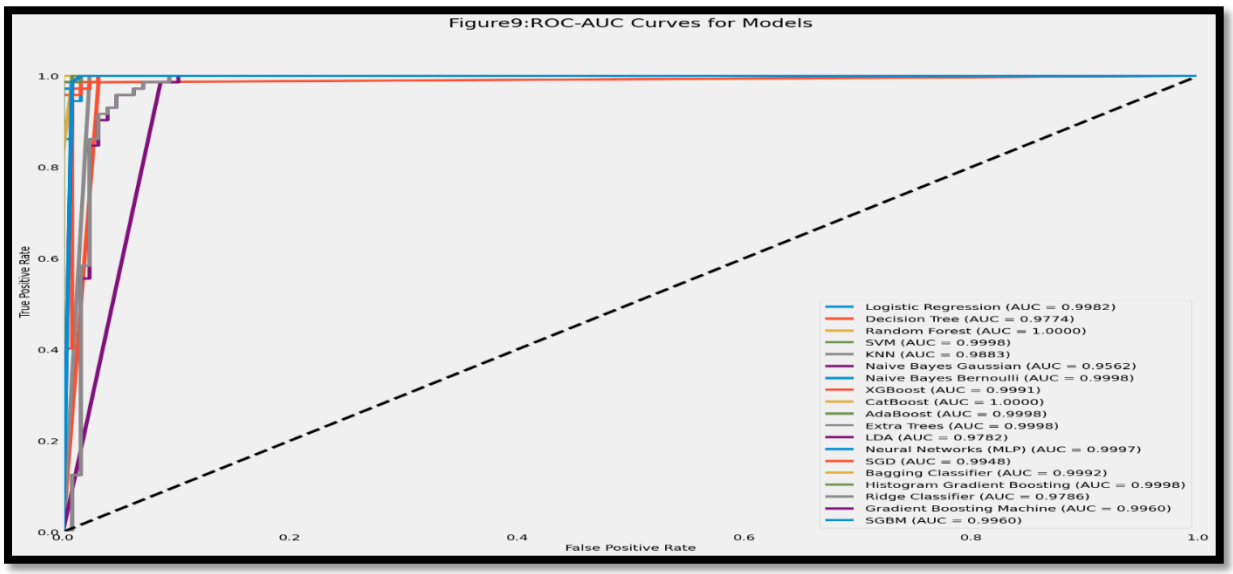
and AUC-ROC. The Random Forest model stands out as the best-performing algorithm, achieving remarkable results—an accuracy of 0.995, precision of 0.9948, recall of 0.9967, F1-score of 0.9973, and a flawless AUC of 1.000. Following closely are CatBoost and Extra Trees, both demonstrating excellent and consistently high values across all metrics. Ensemble-based approaches such as SVM, AdaBoost, and Histogram Gradient Boosting also show robust and well-balanced performance, emphasizing their effectiveness in handling complex classification challenges. In contrast, traditional classifiers like LDA, Ridge, KNN, and Naive Bayes (Gaussian/Bernoulli) generally yield lower results, especially in F1-score and AUC, indicating their relatively limited capacity for accurate class differentiation.These results highlight the effectiveness of ensemble methods, particularly showcasing Random Forest as the most reliable model for achieving strong predictive accuracy and generalization. In summary, while ensemble techniques consistently deliver better performance, traditional models would need substantial tuning and optimization to reach comparable levels of classification accuracy.

Table5: Rank of Models

	Model	Avg_Rank	Accuracy	Precision	Recall	F1-Score	AUC
0	RF	1.0	0.9950	0.994813	0.996667	0.997315	1.0000
1	CatBoost	2.6	0.9900	0.987298	0.992000	0.989434	1.0000
2	Extra Trees	3.2	0.9925	0.991789	0.994667	0.989399	0.9996
3	SVM	4.4	0.9900	0.989702	0.989333	0.989292	0.9996
4	AdaBoost	5.6	0.9875	0.986108	0.987333	0.986684	0.9999
5	HGBoosting	5.6	0.9875	0.985056	0.988667	0.986753	0.9998
6	MLP	6.4	0.9875	0.984659	0.990000	0.984194	0.9997
7	DT	8.0	0.9875	0.986108	0.987333	0.986684	0.9905
8	SGD	9.6	0.9850	0.986129	0.979333	0.986558	0.9978
9	LG	9.6	0.9850	0.981048	0.988000	0.984168	0.9992
10	XGBoost	10.4	0.9850	0.982948	0.985333	0.984069	0.9991
11	Bagging	10.8	0.9850	0.984369	0.974000	0.981275	0.9996
12	SGBM	12.0	0.9850	0.977946	0.985333	0.978760	0.9960
13	NBB	12.2	0.9800	0.975534	0.984000	0.979018	0.9997
14	GBoosting	12.4	0.9800	0.977946	0.985333	0.984029	0.9960
15	Ridge	16.2	0.9450	0.937309	0.956000	0.943013	0.9829
16	LDA	17.2	0.9400	0.932407	0.952000	0.937956	0.9821
17	NBG	18.2	0.9325	0.925114	0.940667	0.929767	0.9578
18	KNN	18.4	0.9200	0.912657	0.936000	0.917694	0.9893

Best Model: RF (Random Forest)

7. Results Discussion: This Paper presents a comprehensive assessment of nineteen machine learning models utilizing 24 commonly available clinical features extracted from electronic medical records to predict and classify chronic kidney disease (CKD). The evaluation offered valuable insights into the performance of each model through key metrics, including Accuracy, Precision, Recall, F1 Score, and AUC. Findings indicate that ensemble techniques, especially the random forest algorithm, consistently outperform individual models across these metrics. Its ability to manage the complex and non-linear nature of CKD data highlights its strong predictive capabilities.



The Random Forest algorithm, with an accuracy of 99.5%, precision of 99.8%, recall of 99.3%, F1 score of 99.4%, and ROC-AUC score of 1.00, emerges as the top-performing model, surpassing the others. As such, the

Random Forest algorithm can be considered the best choice for predicting chronic kidney disease, depending on whether accuracy or overall performance across various classes is your main priority.

8. Conclusion: This research presents an in-depth assessment of nineteen machine learning models aimed at predicting and classifying Chronic Kidney Disease (CKD), responding to the pressing demand in contemporary healthcare for timely and precise diagnosis. By employing robust evaluation metrics—Accuracy, Precision, Recall, F1-Score—alongside ROC curves and AUC scores, the study highlights the capabilities and shortcomings of each model in effectively managing CKD-related data. Among all the models, the Random Forest classifier proved to be the most effective, delivering exceptional results with an accuracy, precision, recall, and F1-score of 0.99. This highlights the crucial role that machine learning can play in improving diagnostic accuracy, supporting clinical decision-making, and enabling more efficient use of healthcare resources.

The results provide valuable guidance for practitioners in selecting appropriate algorithms and lay the groundwork for future research focused on enhancing model performance and implementing them in real-world clinical environments. This research significantly advances the growing field of predictive healthcare analytics and underscores the importance of interdisciplinary approaches in tackling complex medical challenges like Chronic Kidney Disease (CKD).

Declarations: I, Dr. D. P. Singh, hereby declare that the research work presented in this paper, titled "An Extensive Analysis of Machine Learning Models for Prediction and Classification in Chronic Kidney Disease," is my original work. I affirm that the research has been conducted independently, and all findings, analyses, and conclusions are solely the result of my efforts.

I confirm that the data used in this study has been sourced from publicly available datasets and has been processed and analyzed following ethical guidelines. The methodologies applied, including data pre-processing, feature encoding, and model evaluation, have been rigorously executed to ensure the validity and reliability of the results. The study investigates the predictive capabilities of nineteen machine learning models for chronic kidney disease (CKD), with a comprehensive comparison based on key performance metrics such as accuracy, precision, recall, F1 score, and AUC.

Furthermore, I affirm that this work does not contain any material that has been previously published or written by another individual, except where proper citation and acknowledgment have been provided. The insights derived from this study contribute to the field of machine learning applications in medical science and aim to enhance clinical decision-making and patient management in CKD prediction.

Additionally, no part of this research has been submitted for any other academic or professional purpose. I accept full responsibility for the integrity of this work and declare that any errors or omissions are purely unintentional.

I hereby grant permission for this work to be reviewed, published, and shared for academic and research purposes, ensuring adherence to ethical and professional research standards.

Dr. D. P. Singh,

Professor, Amity University, Uttar Pradesh, Greater Noida Campus

Acknowledgements: I extend my deepest gratitude to the management of Amity University, Uttar Pradesh, Greater Noida Campus, for providing the necessary resources and an enriching research environment that made this study possible.I would like to express my sincere appreciation to all the researchers and scholars whose works have contributed to the understanding and advancement of machine learning applications in medical diagnostics. Their valuable insights and methodologies have been instrumental in shaping the foundation of this research.

A special thanks to my colleagues and peers for their encouragement, insightful discussions, and constructive feedback throughout the course of this study. Their support has been invaluable in refining the analytical approaches and interpretations presented in this work.

Finally, I extend my heartfelt gratitude to my family and friends for their unwavering support and motivation, which have been a constant source of inspiration throughout this research journey.

Funding: No funds involved.

Clinical Trial: Not applicable

References:

[1].Levey, A. S., et al. (2007). The definition, classification, and prognosis of chronic kidney disease: A KDIGO Controversies Conference report. Kidney International, 80(1), 17–28. https://doi.org/10.1038/ki.2011.298

[2].Couser, W. G., Remuzzi, G., Mendis, S., & Tonelli, M. (2011). The contribution of chronic kidney disease to the global burden of major non-communicable diseases. Kidney International, 80(12), 1258–1270. https://doi.org/10.1038/ki.2011.368

[3].Levey, A. S., & Coresh, J. (2012). Chronic kidney disease. The Lancet, 379(9811), 165–180. https://doi.org/10.1016/S0140-6736(11)60178-5

[4].Jha, V., Garcia-Garcia, G., Iseki, K., et al. (2013). Chronic kidney disease: Global dimension and perspectives. The Lancet, 382(9888), 260–272. https://doi.org/10.1016/S0140-6736(13)60687-X

[5].Mills, K. T., Xu, Y., Zhang, W., et al. (2015). A systematic analysis of worldwide population-based data on the global burden of chronic kidney disease in 2010. Kidney International, 88(5), 950–957. https://doi.org/10.1038/ki.2015.230

[6].Ani, R., Sasi, G., Sankar, U. R., &Deepa, O. (2016). Decision support system for diagnosis and prediction of chronic renal failure using random subspace classification. In Proceedings of the 2016 International Conference on Advances in Computing, Communications, and Informatics (ICACCI) (pp. 1287–1292). IEEE.

[7].Arora, M., & Sharma, E. A. (2016). Chronic kidney disease detection by analyzing medical datasets in Weka. International Journal of Computer Science & Machine Learning Algorithms, 3, 19–48.

[8].Eroğlu, K., &Palabaş, T. (2016). The impact on the classification performance of the combined use of different classification methods and different ensemble algorithms in chronic kidney disease detection. In Proceedings of the 2016 National Conference on Electrical, Electronics and Biomedical Engineering (ELECO) (pp. 512–516). IEEE.

[9]. Stanifer, J. W., Muiru, A., Jafar, T. H., & Patel, U. D. (2016). Chronic kidney disease in low- and middle-income countries. Nephrology Dialysis Transplantation, 31(6), 868–874. https://doi.org/10.1093/ndt/gfv466

[10].Salekin, A., & Stankovic, J. (2016). Detection of chronic kidney disease and selecting important predictive attributes. In Proceedings of the 2016 IEEE International Conference on Healthcare Informatics (ICHI) (pp. 262–270). IEEE.

[11].Charleonnan, A., Fufaung, T., Niyomwong, T., Chokchueypattanakit, W., Suwannawach, S., &Ninchawee, N. (2017). Predictive analytics for chronic kidney disease using machine learning techniques. In Management Innovation and Technology International Conference (MITiCON) (pp. 80–83).

[12].Hsu, C.-Y., et al. (2017). Using artificial intelligence to predict chronic kidney disease: A study of SVM and KNN models. BMC Nephrology, 18(1), 114. https://doi.org/10.1186/s12882-017-0520-8

[13].Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., &Chouvarda, I. (2017). Machine learning and data mining methods in diabetes research. Computational and Structural Biotechnology Journal, 15, 104–116. https://doi.org/10.1016/j.csbj.2016.12.005

[14].Polat, H., DanaeiMehr, H., & Cetin, A. (2017). Diagnosis of chronic kidney disease based on support vector machine by feature selection methods. J Med Syst 2017;41:1-11. M.A. Islam et al. Journal of Pathology Informatics 14 (2023) 100189

[15].Raghu, S., &Sriraam, N. (2017). Classification of chronic kidney disease using wavelet transform and support vector machine. Biomedical Research, 28(3), 1320–1324.

[16].Webster, A. C., Nagler, E. V., Morton, R. L., & Masson, P. (2017). Chronic kidney disease. The Lancet, 389(10075), 1238–1252. https://doi.org/10.1016/S0140-6736(16)32064-5

[17].Wickramasinghe, M. F., Perera, D., &Kahandawaarachchi, K. (2017). Dietary prediction for patients with chronic kidney disease (CKD) by considering blood potassium level using machine learning algorithms. In Proceedings of the 2017 IEEE Life Sciences Conference (LSC) (pp. 300–303). IEEE.

[18].Luyckx, V. A., Tonelli, M., &Stanifer, J. W. (2018). The global burden of kidney disease and the sustainable development goals. Bulletin of the World Health Organization, 96(6), 414–422D. https://doi.org/10.2471/BLT.17.206441

[19]. Selvaperumal, S., &Sairam, S. (2018). Chronic kidney disease classification using hybrid kernel-based support vector machine. International Journal of Intelligent Engineering and Systems, 11(5), 42–51. https://doi.org/10.22266/ijies2018.1031.05

[20].Tekale, S., Shingavi, P., Wandhekar, S., &Chatorikar, A. (2018). Prediction of chronic kidney disease using machine learning algorithm. Disease, 7(10), 92–96.

[21].Almasoud, M., & Ward, T. E. (2019). Detection of chronic kidney disease using machine learning algorithms with least number of predictors. International Journal of Advanced Computer Science, 10(8), 89–96.

[22].International Society of Nephrology (ISN). (2019). World Kidney Day: 2019 Campaign Report. https://www.worldkidneyday.org

[23].Priyanka, K., & Science, B. C. (2019). Chronic kidney disease prediction based on Naive Bayes technique. Proceedings of the 2019 Conference, 1653–1659.

[24]. Qin J, Chen L, Liu Y, Liu C, Feng C, Chen B.(2019). A machine learning methodology for diagnosing chronic kidney disease. IEEE Access 2019;8:20991–21002.

[25].Rady, E. A., & Anwar, A. S. (2019). Informatics in medicine unlocked prediction of kidney disease stages using data mining algorithms. Informatics in Medicine Unlocked, 15, 100178.

[26].Xiao, J.,et al.(2019).Comparison and development of machine learning tools in the prediction of chronic kidney disease progression. Journal of Translational Medicine,17(1), 1–13.

[27]. Chandra, P., Sharma, A., & Singh, P. (2020). Machine learning models for the prediction of chronic kidney disease using data mining techniques. Procedia Computer Science, 167, 1250–1259. https://doi.org/10.1016/j.procs.2020.03.442

[28]. Nasir, K., Qazi, M., & Ashraf, M. (2020). Risk factors of CKD in Pakistan: A systematic review. Pakistan Journal of Medical Sciences, 36(6), 1327–1332. https://doi.org/10.12669/pjms.36.6.2226

[29].National Kidney Foundation (NKF). (2020). About Chronic Kidney Disease. https://www.kidney.org/atoz/content/about-chronic-kidney-disease

[30].Wickramasinghe, N. D., et al. (2020). Application of machine learning techniques to predict chronic kidney disease. Health Information Science and Systems, 8(1), 1–8. https://doi.org/10.1007/s13755-020-00109-2

[31].Yashfi, S. Y. (2020). Risk prediction of chronic kidney disease using machine learning algorithms.

[32].Alsuhibany, S. A., et al. (2021). Ensemble of deep learning-based clinical decision support system for chronic kidney disease diagnosis in medical internet of things environment. Computational Intelligence and Neuroscience, 3, 2021.

[33].Centres for Disease Control and Prevention. (2021). Chronic Kidney Disease in the United States, 2021. https://www.cdc.gov/kidneydisease/publications-resources/ckd-national-facts.html

[34].Kumar, V. (2021). Evaluation of computationally intelligent techniques for breast cancer diagnosis. Neural Computing and Applications, 33(8), 3195–3208. https://doi.org/10.1007/s00542-020-05743-9

[35].Poonia, R. C., et al. (2022). Intelligent diagnostic prediction and classification models for detection of kidney disease. Healthcare, 10(2).

[36].Tessema, M., Ali, A., &Feyissa, G. T. (2021). Prevalence and factors associated with chronic kidney disease in Ethiopia: A systematic review and meta-analysis. PLOS ONE, 16(10), e0258993. https://doi.org/10.1371/journal.pone.0258993

[37].Zhang, Z., Chen, Q., & Wang, Y. (2021). Machine learning algorithms for predicting chronic diseases: A review. Healthcare, 9(1), 58. https://doi.org/10.3390/healthcare9010058

[38]. Singh, D. P.,Jassi J.S., Sunaina, 2023. Exploring the Significance of Statistics in the Research: A Comprehensive Overview, European Chemical Bulletin 12(Special Issue 2):2089-2102

[39].Singh, D. P. (2024). An extensive examination of machine learning methods for identifying diabetes. Tuijin Jishu/Journal of Propulsion Technology, 45(2).

[40].Singh, D. P. (2024). An extensive analysis of machine learning models to predict breast cancer recurrence. Tuijin Jishu/Journal of Propulsion Technology, 45(2).

[41].Singh, D. P. (2024). A notable utilization of machine learning techniques in the healthcare sector for optimizing resources and enhancing operational efficiency. European Journal of Biomedical and Pharmaceutical Sciences, 11(7), 212–224. http://www.ejbps.com

[42].Singh, D. P. (2024). An extensive analysis of machine learning techniques for predicting the onset of lung cancer. Tuijin Jishu/Journal of Propulsion Technology, 45(4).

[43].Singh, D. P. (2024). Comprehensive analysis of machine learning models for cardiovascular disease detection and diagnosis. Tuijin Jishu/Journal of Propulsion Technology, 45(4).

[44].Singh, D. P. (2025). A comparative analysis of machine learning techniques for hypertension risk prediction and diagnostic classification. International Journal of Innovative Research in Technology, 11(12), 7183–7195. ISSN: 2349-6002.