https://doi.org/ 10.33472/AFJBS.6.5.2024. 5928-5943

African Journal of Biological Sciences

Analyzing Chronic Kidney Disease Performance Using Machine Learning Techniques

Harwinder Singh Sohal*, Kamal Malik *

Research Scholar, Professor

Department of Computer Science & Engineering CT University, Punjab, India

harwindersohal23@gmail.com, professorkamalmalik@gmail.com

Article History
Volume 6, Issue 5, 2024
Received: 25 May 2024
Accepted: 02 Jun 2024
doi: *10.33472/AFJBS.6.5.2024. 5928-5943*

Abstract

A serious condition that can last a lifetime, Chronic Kidney Disease (CKD) is brought on by either impaired kidney function or kidney cancer. It is possible to stop or limit the advancement of this chronic illness to the point when a patient's sole options for survival are dialysis or surgery. Chronic kidney disease (CKD) is caused by illnesses that impair and diminish kidneys' ability to maintain and human health. Consequently, kidney disease must be taken seriously from the very beginning. The chance that this will occur can be raised with early detection and suitable therapy. This study has investigated the potential of a number of machine learning techniques, including eXtreme Gradient Boosting (XGBoost), Adaptive Boosting (AdaBoost), K-Nearest Neighbors (KNN), Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM) and Logistic Regression (LR) for early CKD diagnosis. The performance of every classification algorithm was encouraging. The Decision Tree and Support Vector Machine achieved an accuracy of 98.61% and 97.22% for all measures respectively outperforming all other applied techniques.

KEYWORD: Machine Learning; Chronic Kidney Disease; Random Forest; Decision Tree; Logistic Regression; Glomerular Filtration Rate.

1. Introduction

A pair of kidneys are an essential organ for the body's proper operation. Its job is to filter blood, eliminate waste, manage the body's fluid balance and control the production of urine. The term "Chronic Kidney Disease" (CKD) refers to a disorder that affects kidney function and makes it harder for the kidneys to operate properly, producing more waste in the blood that over time makes the human body sick (Abdulah, 2021). A widespread ailment known as Chronic Kidney Disease refers to a long-term decline of kidney function and affects millions of people worldwide. Chronic Kidney Disease which is becoming increasingly prevalent and causes a decline in kidney function, is a condition brought on by the modernization of people's daily lives and the extension of life spans in society (Dilli Arasu S, 2017). When chronic kidney disease first develops, one may have simply a few symptoms or indicators. It's possible that chronic kidney disease won't show symptoms until kidney function is seriously compromised. Without mechanical filtering (dialysis) or a kidney transplant, end-stage kidney failure from Chronic Kidney Disease is fatal (Janani, 2021). The absence of symptoms in the early stages of CKD, a common ailment that takes time to emerge, causes people with the condition to undervalue and neglect their condition. The pathology manifests at a fairly advanced level as a result of the increasing loss of function. As an asymptomatic condition, its complications are difficult to identify and can be mistaken for those of other conditions. The majority of the time, CKD is discovered with kidney failure, which often necessitates drastic procedures like kidney transplantation or if those fail, it causes death. Only skilled nephrologists and urologists are able to identify CKD at its early stages utilizing the disease's history, symptoms and laboratory tests (Ana Pinto, 2020).

1.1. Symptoms

Chronic Kidney Disease frequently starts out with minor symptoms before moving on to more severe ones. Effective treatment and early detection are the primary cures for lowering the rate of mortality as chronic kidney disease steadily progresses. Diabetes mellitus, high blood pressure, autoimmune disorders, systemic (body-wide) infections, urinary tract infections, urinary stones, obstructions of the lower urinary tract and toxic medicines are some of its symptoms (Abdulah, 2021). These symptoms may affect both their capacity to work and quality of life. In addition, it causes nerve damage and other chronic ailments like high blood pressure, anemia and weak bones as a result of poor nutritional health. While this is going on, kidney illness increases a

patient's chance of developing blood and heart disorders. Even other chronic diseases like diabetes, high blood pressure and other problems are brought on by chronic kidney disease. People in high risk groups include inherited conditions like diabetes and hypertension. Through early discovery and appropriate treatment, chronic kidney disease can be cured. However, if the condition is allowed to continue, kidney failure may result (Dilli Arasu S, 2017). Because kidney disease can cause a considerable reduction in kidney function without exhibiting any symptoms, it is known as a silent killer.

1.2. Stages of CKD

Chronic Kidney Disease is a serious health issue that impacts people worldwide over the globe. Getting the wrong CKD diagnosis might have catastrophic consequences, which impact individuals who are unable to be treated. There are five stages of CKD with fifth stage being the deadliest because the kidneys cannot perform the majority of their tasks at this point. The ability of the kidneys to remove waste and extra fluid from the circulation affects the stages of kidney disease. Even in the early stages of kidney disease, the kidneys can still remove waste from the blood. The kidneys must work harder and may even stop working altogether in the later stages of the disease. It is challenging to determine each patient's stage of CKD especially in the early stages. The best test to assess kidney health and gauge a person's degree of Chronic Kidney Disease is the Glomerular Filtration Rate (GFR) (Janani 2021). If GFR is high and there are no symptoms, kidney function is regarded as normal and the kidneys are healthy. Stage 1 kidney damage includes normal kidney function and a GFR of greater than 90. In stage 2, GFR is modestly reduced and ranges from 89 to 60. The third stage of kidney disease is thought to have moderate loss of kidney function, which occurs when GFR falls from 59 to 30. Stage 4 is only defined as a severe GFR reduction when the GFR is between 29 and 15. The final stage of chronic kidney disease is kidney failure. When GFR is less than 15, it is referred to as the 5th stage or ESRD or End Stage Kidney Disease (Madhur Bhatt, 2019). The most important measure of glomerular filtration rate (GFR) reliable test to diagnose the stage of chronic kidney disease and kidney function. Blood creatinine levels, age, gender and other characteristics can all be used to compute it. A prolonged decrease in Glomerular Filtration Rate (GFR) or indications of structural and functional abnormalities in the kidneys are considered to be signs of chronic kidney disease (CKD) (Pal D. C., 2021)

2. Literature Review

Asif Salekin (2016) evaluated a method using a dataset of 400 people, 250 of whom have CKD and get a detection accuracy of 99.3% and a root mean square error of 0.1084. Başar et al.(2016) used Adaboost ensemble learning method to diagnose chronic kidney disorders. Classifiers built on decision trees were employed in the diagnosis. The kappa, mean absolute error (MAE), root mean squared error (RMSE) and area under curve (AUC) metrics were used to assess the classification performance. Adaboost ensemble learning algorithm offers greater classification performance than individual classification, according to performance assessments. Boukenze et al. (2017) used several machine learning methods in their research to predict kidney illness including Support Vector Machine (SVM), Multilayer Perceptron (MLP), Decision Tree (C4.5), Bayesian Network (BN) and K-Nearest Neighbor (K-NN). According to the experimental findings MLP and C4.5 have the highest accuracy rates. However, C4.5 seems to be the most effective when compared to the Receiver Operating Characteristic (ROC) curve and demonstrated superior performance to MLP on a number of metrics by having the lowest error rate and fastest execution time. Chimwayi et al. (2017) used the neuro-fuzzy algorithm to build a model for predicting risk in CKD taking into account all of the factors that can cause the disease's symptoms and progression assess a patient's CKD risk. Neuro-fuzzy predictions have a 97% accuracy rate. Amirgaliyev et al. (2018) proposed a noninvasive, inexpensive and efficient computerized classification method. The results demonstrated that a classifier with adequate implementation can achieve an overall performance of 94.602%. The SVM classifier's linear kernel's sensitivity value is 93.100%. Drall et al. (2018) used two classification algorithms KNN and Naive Bayes as well as five CKD related variables out of a total of 25 attributes to predict patient's CKD status. The accuracy of the KNN classifier's prediction of chronic kidney disease was 100%, compared to the Naive Bayes Classifier's prediction accuracy of 96.25%. KNN thus outperforms Nave Bayes with excellent accuracy. Pujianto et al. (2018) investigated the application of the Support Vector Machine and K-Means algorithms for clustering and classification respectively. According to the findings of the study, the classification procedure using two clusters yields the maximum accuracy which is 100% for all kernel functions. Adnan et al. (2019) used MATLAB software (fuzzy logic toolbox) to apply the predictive model in CKD detection. Clinical test results from 70 patients were used as a set of data in their investigation. The outcome reveals that CKD was identified in 47 of the 70 patients. Bhatt and Kasbe (2019) conducted a survey regarding the individuals' expert system for the diagnosis of

chronic kidney disease and suggested how much work needs to be done before establishing an expert system for diagnosing chronic kidney illness utilizing reliable data from the UCI dataset machine learning repository and strong reasoning from fuzzy logic. Devika et al. (2019) described specialized classifiers to predict chronic kidney disease and it concluded with a performance comparison analysis. Authors evaluated the accuracy, precision and execution time of the Naive Bayes, K-Nearest Neighbor (KNN) and Random Forest classifiers for CKD prediction. The final finding of the research is that Random Forest classifier performs better than Naive Bayes and KNN. Das et al. (2020) used eight benchmark datasets, this suggested model is evaluated for performance and contrasted with alternative models. To verify the accuracy of the results, statistical methods like Friedman and Holm-Bonferroni were used. Hamedan (2020) created a fuzzy logic-based expert system for the detection and compared the system's performance on the original and noisy datasets after adding noisy data to our dataset. The final system's precision, sensitivity and specificity were 92.13%, 95.37% and 88.88% respectively. Islam et al. (2020) suggested that risk factors for chronic kidney disease (CKD) can be predicted. The risk of eliminating chronic kidney disease (CKD) is significantly reduced by risk factor predictions. The greatest results were obtained by classifying each and every approach and using algorithms to anticipate risk variables. The Random Forest algorithm helped the proposed model to achieve a high accuracy rate of 98.8858 %. Yashfi et al. (2020) examined the records of CKD patients and put forth a system that would allow for the prediction of CKD risk. Here, an online data set of 400 patients from the UCI Machine Learning Repository and 55 real life dataset from Khulna City Medical College were combined. Author's used Random Forest and ANN after 10-fold CV to train the data. Random Forest algorithm achieved an accuracy of 97.12% while ANN accuracy was 94.5%. Almustafa (2021) presented a prediction model based on feature selection to predict CKD effectively. The methods were implemented utilizing Random Tree, Decision Table, K-Nearest Neighbor, J48, Stochastic Gradient Descent and Naïve Bayes classifiers. The decision table and J48 classifiers performed better than the other classifiers, as evidenced by their respective 99% accuracy rates, ROC values of 0.999 and 0.992, MAEs of 0.0225 and 0.1815 and RMSEs of 0.0807 and 0.2507. Chicco et al. (2021) analyzed the medical records of patients with CKD and CVD using machine learning. They made predictions regarding whether patients will experience severe CKD with a high mean Matthews correlation coefficient (MCC) of C0:499 and in the latter case a mean MCC of was C0:469. Damodara and Thakur (2021) used the adaptive neurofuzzy logic system (ANFIS) to show how kidney illnesses can be predicted

early. ANFIS CKD stage prediction model built in a Matlab gave accuracy of 94% when comparing real output to estimated output. Emon et al. (2021) used eight machine learning classifiers, including the Logistic Regression, Naïve Bayes, Multilayer Perceptron, Stochastic Gradient Descent, Adaptive Boosting, Bagging, Decision Tree and Random Forest classifiers to determine Chronic Kidney Disease. The Random Forest (RF) algorithm has the best ROC value and 99% accuracy. Ilyas et al. (2021) compared two algorithms J48 and Random Forest to predict the different stages of CKD, they discovered that J48 properly classifies cases at a ratio of 85.5%, compared to Random Forest's ratio of 78.25%. In contrast, J48 takes 0.03 s and Random forest 0.28 s to complete the task. Yadav and Pal (2021) examined neural networks for chronic kidney disease with the aid of features reduction and pertinent approaches. Author's developed a training model using 300 (75%) cases of features related to chronic kidney illness and tested it using 100 (25%) instances and attained an accuracy of 99.98% with the lowest error rate. Murugesan et al. (2022) used fuzzy and adaptive neural fuzzy inference systems to identify chronic kidney disease. The reliability indicator was examined using the aforementioned method and it was found that 93.75% of the findings from the fuzzy inference system were accurately classified. Poonia et al. (2022) provided a good feature-based prediction model using a variety of machine learning algorithms including the K-Nearest Neighbors Algorithm (KNN), Artificial Neural Networks (ANN), Support Vector Machines (SVM), Naive Bayes (NB). RFE and Chi-Square test feature-selection techniques were also used. The research discovered that a prediction model based on logistic regression with the best features selected using the Chi-Square method had the maximum accuracy of 98.75%. Srivastava et al. (2022) proposed a performance tuning nested approach (ranking weighted ensemble) with 98.75% accuracy, 100% sensitivity, 96.55% specificity and 99.03% f1score that considers adjusting hyper-parameters as well as determining the appropriate weights to join ensembles. Yan et al. (2022) created prediction models for the risk assessment of CKD. XGBoost, Random Forest, Naive Bayes, Support Vector Machines and Multivariate Logistic Regression techniques were used. They compared each model's accuracy, precision, recall, F1 score and area under the receiver operating curve (AUC). The XGBoost model and the random forest model both showed good prediction effects, with the XGBoost model showing the highest results (accuracy 0.9088, precision 0.9175, recall 0.8244, F1 score 0.8868 and AUC 0.8244). Salman and Gupta (2023) suggested a hybrid framework that incorporates Logistic Regression for feature extraction and Random Forest to categorize the

diseases. The suggested framework is simulated using Python. The results are analyzed using a variety of criteria, including accuracy, precision and recall.

3. Materials and Method

The present work focus on applying machine learning techniques to predict the Chronic Kidney Disease by considering accuracy as an important factor. The proposed model consists of two main approaches. In the first approach essential features from the Chronic Kidney Disease dataset is selected by using feature selection methods. In the second approach apply different machine learning techniques on the selected features to predict the Chronic Kidney Disease. Various steps involved in the implementation of proposed model is described in Fig. 1.

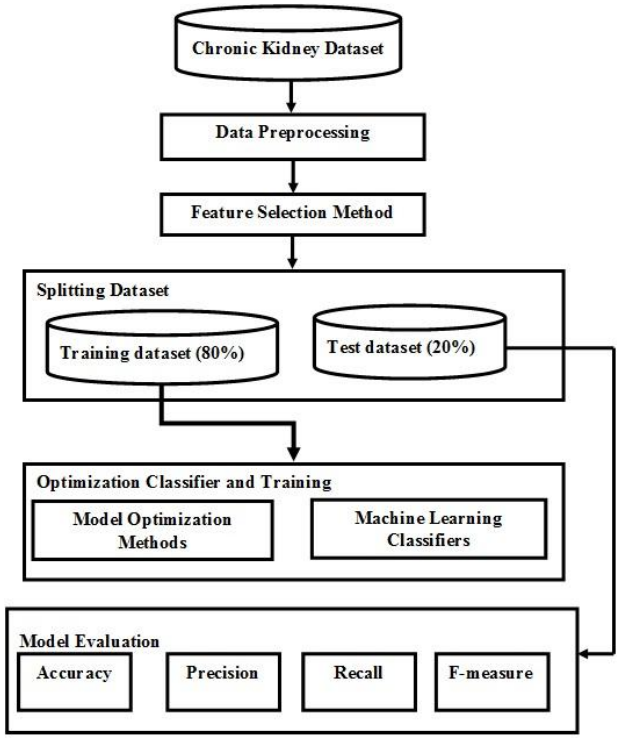

Fig. 1. Procedure for implementing the proposed work

3.1. Data source and description

For analysis, a dataset of 400 patients was chosen from the UCI Machine Learning Repository each with 25 attributes. There were 250 kidney disease patient records in the dataset and 150 medical records for healthy individuals. This dataset contains health information for several age groups. There are 50 records of individuals under 30 and 55 records of individuals over 70 years old. Those in the remaining records range in age from 31 to 69. Kidney illness can affect anyone

at any age, according to the numerous research. Details of the different kidney disease-related characteristics are displayed in Table I.

Table I: Description of the Kidney Disease-Related Attributes

Attribute	Description	Type	Information
age	Patient's age	Numerical	Years (0-90)
bp	Blood pressure of the patient	Numerical	mm/Hg (0-180)
sg	Ratio of the density of urine	Nominal	1.005-1.025
al	Ratio of the albumin level in blood	Nominal	0,1,2,3,4,5
su	Sugar level of the patient	Nominal	0,1,2,3,4,5
sc	Serum creatinine level in the blood	Numerical	mgs/dl (0-76)
sod	Sodium level in the blood	Numerical	mEq/L (0-163)
pot	Potassium level in the blood	Numerical	mEq/L (0-47)
wc	White blood cell count of the patient	Numerical	cells/cumm (0-26,400)
rc	Red blood cell count of the patient	Numerical	cells/cumm (0-8)
hemo	Hemoglobin level in the blood	Numerical	gms (0-17.8)
pcv	Packed cell volume in the blood	Numerical	mEq/L (0-54)
bu	Blood urea level of the patient	Numerical	mgs/dl (0-391)
bgr	Blood glucose random count	Numerical	mgs/dl (0-490)
cad	Coronary artery disease or not	Nominal	yes, no
appet	Patient's appetite	Nominal	good, poor
rbc	Patients' red blood cell counts	Nominal	normal, abnormal
pc	pus cell count of patient	Nominal	normal, abnormal
pcc	Pus cell clumps in the blood	Nominal	present, notpresent
htn	Patient have hypertension on not	Nominal	yes, no
dm	Patient have diabetes or not	Nominal	yes, no
pe	Patient have pedal edema or not	Nominal	yes, no
ane	Patient have anemia or not	Nominal	yes, no
ba	Presence of bacteria in the blood	Nominal	present, not present
classification	Healthy or Unhealthy people	Nominal	ckd ,not ckd

3.2. Proposed Algorithm

Procedure: The suggested method for identifying kidney disease.

Input: Dataset containing records on kidney disease.

Output: The prediction models' ability to identify kidney disease.

Stage 1 Preprocessing of Dataset

The values in the "rbc" and "pc" columns of the original dataset are normal, absent and empty. There are 150 and 120 entries in the "rbc" and "pc" columns respectively that are blank. The columns labeled "pcc" and "ba" in this dataset include values for "present" and "not present." The values "yes" and "no" are present in the columns labeled "cad," "pe," "htn," "dm," and "ane." The terms "appet" and "good" are also included in this dataset. Thus, in order to get accurate

findings, preprocessing this dataset is required. Essentially, NaN is used in place of the empty values. The following is how conversion of nominal values into binary values us performed:

1. "Normal" and "abnormal" nominal values are swapped out for 1 and 0, respectively, in the columns labeled "rbc" and "pc."
2. The values for "present" and "nonpresent" are changed to 1 and 0, respectively, in the columns labeled "pcc" and "ba."
3. 1 and 0 are used in place of the values "yes" and "no" in the columns labeled "htn," "pe," "ane," "dm," and "cad."
4. Lastly, "good" and "poor" are changed to 1 and 0, respectively, in the "appet" column.

Stage 2 Examine how various features relate to one another

This stage involves determining the connection between the input and target attributes. Proposed research revealed a weak relationship between "pot" and "ba" and the objective attribute.

Stage 3 Split the dataset

Using an 80:20 ratios the dataset is partitioned into training and testing datasets in this stage. This indicates that 20% of data are utilized for testing and 80% of data are used for training.

Stage 4 Configure the machine learning techniques' settings

In this step, machine learning methods are used to the processed features of the kidney disease dataset in order to construct prediction models. In order to construct the prediction model, implement Machine Learning techniques such as Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), eXtreme Gradient Boosting (XGBoost) and Adaptive Boosting (AdaBoost) and 10-fold cross validation is to be used.

Stage 5 Feature selection

A machine learning model may become overfitted if there are too many characteristics or too many features that are the same. Reducing the number of features in the dataset is essential to prevent this. Feature selection in machine learning refers to the process of reducing the number of features in a dataset. Reducing the number of input attributes lowers the cost of building the machine learning model and, in certain situations, may even improve the accuracy of the model.

Stage 6 Construct the prediction model

In order to create distinct prediction models, 10-fold cross validation is used in this step, along with a variety of machine learning techniques and the chosen characteristics.

Stage 7 Performance Evaluation

Lastly, a comparison is made between the performance of prediction models using all features and just some features.

4. Results and Analysis

Various performance measures were used to compare and analyze the output of all machine learning techniques. Python 3.11.4 was used for the studies and it may be used with either the Jupyter Notebook web application or Google Collab. Numerous libraries were used in the implementation and Sciket-learn which is a crucial library for building ML models was one of them. In the present work many confusion matrix performance parameters are taken into account. The major purpose of this is to design and create a standard predictive model in terms of algorithm, which will be beneficial for any patients who have chronic kidney disease and will be used for future predictions by both nephrologists and the medical sector. In the current investigation, dataset of 400 individuals with 25 attributes collected from UCI machine learning repository's and was subjected to seven Machine learning techniques. The best accurate algorithm for categorizing the severity stage of CKD has been determined by comparing the outcomes of the algorithms that were addressed. To move further with the prediction, the entire dataset must be divided into the training subset and the testing subset. The testing subset is used for testing, whereas the training subset is utilized for training. The training subset's results are known, but the testing subset's prediction result or output are unknown. The CKD dataset was divided at 80:20 ratios. Tenfold cross-validation has been used for these models' testing and training. The dataset is randomly divided into ten equal-sized sets for tenfold cross-validation. After that, the models were trained using ten to one fold and tests were conducted using the last fold. For every fold, the procedure is repeated. The model's performance was assessed by displaying the correctly and erroneously identified classes in the confusion matrix. The seven machine learning techniques used for the assessment of Chronic Kidney Disease are: Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), K-

Nearest Neighbors (KNN), eXtreme Gradient Boosting (XGBoost) and Adaptive Boosting (AdaBoost).

4.1. Performance Measures

Confusion matrix: A table that is used to describe how well a learning algorithm performs after computing performance measures. Confusion matrix displays the classification model's correct and wrong predictions in relation to the dataset's actual results or intended value. There are two predicted classes in present work: "CKD" and "not CKD." If someone forecasting the presence of a disease, "CKD" would indicate that the person has the chronic kidney disease while "not CKD" would indicate that they do not. The experiment employed the metrics listed below to determine performance:

True Positive (TP): Those events are that are appropriately identified as positive outputs and indicate that a person has CKD.

True Negative (TN): The expected indicates accurately classified negative outputs as negative outputs. They have chronic kidney disease even when the prediction is not CKD.

False Positive (FP): Individuals who are diagnosed with CKD but do not actually have it The forecast shows negative occurrences that were mistakenly categorized as positive outcomes.

False Negative (FN): The person actually has chronic kidney disease despite the prediction that they did not. It denotes instances of positive output that were mistakenly categorized as negative. The number of test records divided by the total number of successfully classified records yields the accuracy. Precision is the ratio of True Positive (TP) records to the total number of TP records in a given class. True positives and false negatives are the two categories of recall. The recall ratio measures how many records were correctly classified out of all the records in a class. A range of performance indicators were employed by to evaluate the efficacy of machine learning techniques. Parameters like precision, recall, F-measure and accuracy performance measures were employed to assess and contrast the effectiveness of the suggested prediction models. To calculate the accuracy, precision, recall and F-measure the following formulas were applied.

Accuracy: A popular metric for classification methods is accuracy. It is the percentage of samples to all samples that were accurately classified. The ratio of correct forecasts to total predictions is known as accuracy. This ratio assesses how well the model classified true positives

as positive and true negatives as negative. Regarding this matter, it provides data regarding the proportion of patients among all patients who receive an accurate diagnosis of chronic kidney disease and those who do not.

$$\textbf{Accuracy} = (TP+TN)/(TP+TN+FP+FN)$$

Where FN is the number of records that were not classified as having a kidney disease, TN is the number of images that were not assigned to the correct kidney disease class, TP represents the total number of records that were accurately assigned to the kidney disease class and FP is the number of records incorrectly classified as having kidney disease.

Precision: The definition of precision is the proportion of affirmative identifications that were actually correct. It measures the percentage of expectedly positive values that really turn out to be positive. Here, it represents the percentage of people who have been diagnosed with chronic kidney disease.

$$\textbf{Precision} = TP/(TP+FP)$$

Recall: It is sometimes referred to as true positive rate (TPR), hit rate and sensitivity. The percentage of correctly classified positive examples relative to all positive instances is displayed. Recall quantifies the number of true positives that the model correctly labels as positive. In cases where FN (such as a medical diagnosis) carries a significant expense, this is the criterion to take into account. The quantity of CKD patients who could accurately predict is measured by the recall metric.

$$\textbf{Recall} = TP/(TP+FN)$$

F-Measure: It gauges the test's accuracy. It's the harmonic mean of recall and precision.

$$\textbf{F-Measure} = 2*(Recall*Precision) / (Recall + Precision)$$

4.2. Discussions

Table II lists the experimental findings for each developed model. The outcomes demonstrate that the suggested methodology is workable. The key features that are crucial for the diagnosis of Chronic Kidney Disease were extracted using the Information Gain feature selection technique. It was discovered that particular parameters with a higher information gain such as specific gravity, albumin, serum creatinine, haemoglobin, packed cell volume, red blood cell count and hypertension were more significant than the other features. The performance of machine learning

techniques was improved by KNN Imputation compared to scenarios where random imputation removing rows with missing data or mean and mode imputation were utilized. By substituting the missing values with the mean of the k nearest neighbors, KNN Imputation addressed the missing data. The experiment involved imputer optimization and it was discovered that 5 was the most ideal value for k for the CKD dataset. According to the evaluation's findings, each model performs admirably in terms of accurately identifying CKD. Table II shows that, when compared to the other models Decision Tree (98.61%) and Support Vector Machine (97.22%) performed the best with a greater level of accuracy also shown graphically in Fig. 2.

TABLE II: Evaluation Metrics

Machine Learning Algorithms	Precision	Recall	F-Measure	Accuracy
Logistic Regression (LR)	*93.33%*	*96.55%*	*95%*	*95.83%*
Decision Tree (DT)	*100%*	*96.55%*	*98%*	*98.61%*
Random Forest (RF)	*93.33%*	*96.55%*	*96%*	*95.83%*
Support Vector Machine (SVM)	*93.55%*	*100%*	*97%*	*97.22%*
K-Nearest Neighbors (KNN)	*90.63%*	*100%*	*95%*	*95.83%*
eXtreme Gradient Boosting (XGBoost)	*87.50%*	*96.55%*	*93%*	*93.06%*
Adaptive Boosting(AdaBoost)	*93.33%*	*96.55%*	*96%*	*95.83%*

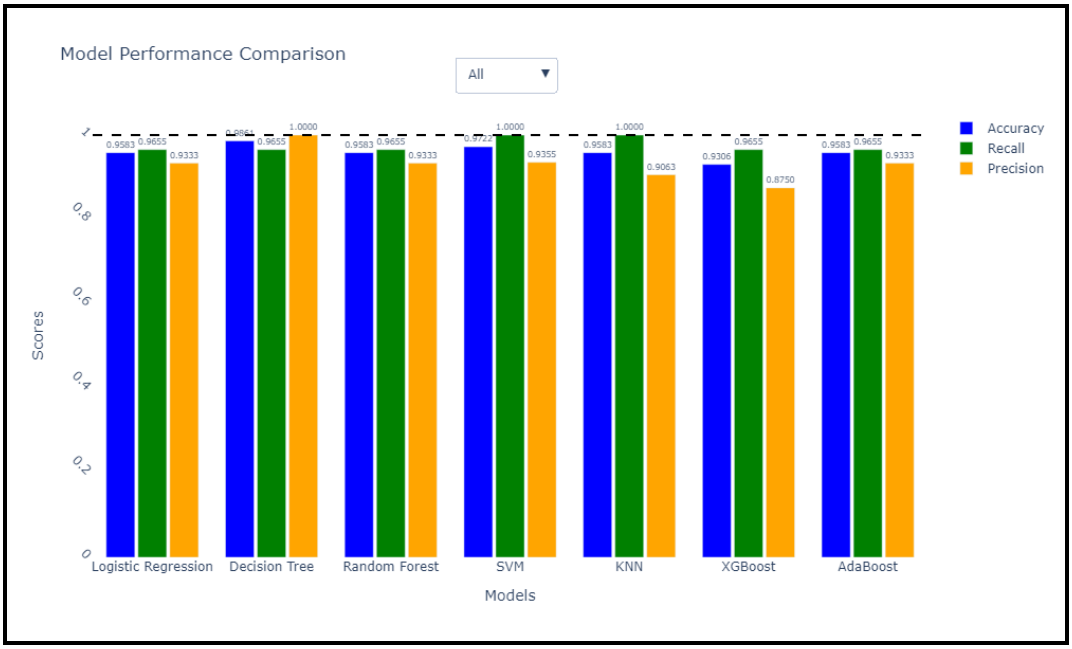

Fig. 2. Performance Analysis of Machine Learning Algorithms

5. Conclusions and Future Work

This paper presents a machine learning model for the early identification of chronic kidney disease. In order to diagnose chronic kidney disease earlier, this paper offered several machine learning techniques. To validate the machine learning-based detection algorithms, the original CKD dataset was preprocessed. Then the most prominent features were found using PCA, which allowed for the detection of CKD. Following the construction of the models employing CKD patients, the input parameters that were previously described are then used to train and validate the model. Research has been conducted on the correlations among various elements to reduce the quantity of features and remove unnecessary data. In terms of predicting chronic kidney disease, hemoglobin, albumin, packed red blood cell count, hypertension, albumin and specific gravity were found to have the largest effects when a filter feature selection approach was applied to the remaining attributes. A set of features is supplied into classification algorithms. The comparison analysis is estimated using many metrics such as classification accuracy, f-measure, recall and precision. The accuracy of such algorithms was the primary criterion that was utilized in evaluating their overall performance. The advantage of this strategy is that the prediction process is faster, allowing physicians to treat CKD patients as soon as possible and classify a larger number of patients in a shorter amount of time. It has been suggested to work with larger datasets in the future or compare the outcomes of this dataset with a different dataset with the same because the dataset utilized in this paper is small, consisting of just 400 samples.

Additionally, using the appropriate information, attempts have been made to predict whether a person with this syndrome has a higher likelihood of chronic risk factors such diabetes, hypertension, and a family history of renal failure in order to assist reduce the incidence of CKD.

6. References

1. Abdulah, N. S. (2021). Diagnosis of Chronic Kidney Disease Using Hybrid Algorithm. 1st International & 4th Local Conference for Pure Science (ICPS-2021), 1-10.
2. Ain Nazirah Binti Mohd Adnan, N. I. (2019). Fuzzy Logic An Application to Detect Chronic Kidney Disease And Failure. Open International Journal of Informatics (OIJI), 118-126.
3. Almustafa, K. M. (2021). Prediction of chronic kidney disease using different classification algorithms. Khaled Mohamad Almustafa, 1-7.
4. Ana Pinto, D. F. (2020). Data Mining to Predict Early Stage Chronic Kidney Disease. International Workshop on Healthcare Open Data, Intelligence and Interoperability (HODII), 562-567.
5. Asif Salekin, J. S. (2016). Detection of Chronic Kidney Disease and Selecting Important Predictive Attributes. IEEE International Conference on Healthcare Informatics, 262-270.
6. Basma Boukenze, A. H. (2017). Predicting Chronic Kidney Failure Disease Using Data Mining Techniques. Advances in Ubiquitous Networking 2, Lecture Notes,in Electrical Engineering 397, DOI 10.1007/978-981-10-1627-1_55, 701-712.
7. Davide Chicco, C. A. (2021). A Machine Learning Analysis of Health Records of Patients With Chronic Kidney Disease at Risk of Cardiovascular Disease. IEEE Access, 165132-165144.
8. Devika R, S. V. (2019). Comparative Study of Classifier for Chronic Kidney Disease prediction using Naive Bayes, KNN and Random Forest. Proceedings of the Third International Conference on Computing Methodologies and Communication (ICCMC 2019), 679-684.
9. Dilli Arasu S, D. R. (2017). Review of Chronic Kidney Disease based on Data Mining Techniques. International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 23, 13498-13505.
10. Farahnaz Hamedan, A. O. (2020). Clinical decision support system to predict chronic kidney disease: A fuzzy. International Journal of Medical Informatics, 1-9.
11. Hamida Ilyas, S. A. (2021). Chronic kidney disease diagnosis using decision tree algorithms. BMC Nephrology, 3-11.
12. Himansu Das, B. N. (2020). Medical disease analysis using Neuro-Fuzzy with Feature Extraction Model for classification. Informatics in Medicine Unlocked 1-14.
13. Janani J, S. R. (2021). Diagnosing Chronic Kidney Disease Using Hybrid Machine Learning Techniques. Turkish Journal of Computer and Mathematics Education Vol.12 No.13, 6383-6390.
14. Kerina Blessmore Chimwayi, N. H. (2017). Risk Level Prediction of Chronic Kidney Disease Using Neuro Fuzzy and Hierarchical Clustering Algorithm. International Journal of Multimedia and Ubiquitous Engineering, 23-36.

15. Komal Damodara, A. T. (2021). Adaptive Neuro Fuzzy Inference System based Prediction of Chronic Kidney Disease. 7th International Conference on Advanced Computing & Communication Systems (ICACCS), 973-976.
16. Madhur Bhatt, T. K. (2019). A Survey on Chronic Kidney Disease Diagnosis using Fuzzy Logic. IEEE International Symposium on Smart Electronic Systems, 252-256.
17. Md. Ashiqul Islam, S. A. (2020). Risk Factor Prediction of Chronic Kidney Disease based on Machine Learning Algorithms. Third International Conference on Intelligent Sustainable Systems [ICISS 2020], 952-957.
18. Merve Doğruyol Başar, P. S. (2016). Detection of Chronic Kidney Disease by Using Adaboost Ensemble Learning Approach. IEEE, 1-4.
19. Minhaz Uddin Emon, A. M. (2021). Performance Analysis of Chronic Kidney Disease through Machine Learning Approaches. Proceedings of the Sixth International Conference on Inventive Computation Technologies [ICICT 2021], IEEE, 713-719.
20. Murugesan G., T. I. (2022). Fuzzy Logic-Based Systems for the Diagnosis of Chronic Kidney Disease. BioMed Research International, Hindawi ,Article ID 2653665, 1-15.
21. Pal, D. C. (2021). Performance based Evaluation of Algorithmson Chronic Kidney Disease using Hybrid Ensemble Model in Machine Learning. Biomedical & Pharmacology Journal, September 2021. Vol. 14(3), 1633-1645.
22. Rahama Salman, S. G. (2023). Hybrid Machine Learning Model for Chronic Disease Prediction. International Journal of Intelligent Systems and Applications in Engineering, 808–816.
23. Ramesh Chandra Poonia, M. K. (2022). Intelligent Diagnostic Prediction and Classification Models for Detection of Kidney Disease. https://www.mdpi.com/journal/healthcare, 2-19.
24. Shanila Yunus Yashfi, M. A. (2020). Risk Prediction Of Chronic Kidney Disease Using Machine Learning Algorithms. 11th ICCCNT, IEEE - 49239, 1-5.
25. Sujata Drall, G. S. (2018). Chronic Kidney Disease Prediction Using Machine Learning: A New Approach. International Journal of Management, Technology And Engineering, ISSN NO: 2249-7455, 278-287.
26. Swapnita Srivastava, R. K. (2022). An Ensemble Learning Approach For Chronic Kidney Disease Clasificación. Journal of Pharmaceutical Negative Results, Volume 13, Special Issue 10, 2401-2409.
27. Utomo Pujianto, N. A. (2018). Support Vector Machine with Purified K-Means Clusters for Chronic Kidney Disease Detection. The 2nd East Indonesia Conference on Computer and Information Technology (EIConCIT), 56-60.
28. Xiaoqian Yan, X. L. (2022). Establishment and Evaluation of Artificial Intelligence-Based Prediction Models for Chronic Kidney Disease under the Background of Big Data. Hindawi, Evidence-Based Complementary and Alternative Medicine,Volume 2022, Article ID 6561721, 1-7.
29. Yedilkhan Amirgaliyev, S. S. (2018). Analysis of Chronic Kidney Disease Dataset by Applying Machine Learning Methods. IEEE, 1-4.