Kafrelsheikh Journal of Information Sciences ISSN (Online): 2535-1478, ISSN (Print): 2537-0677 Volume 4, Issue 2, 2023, PP. 1–6
Date of publication 25 November 2023

Chronic Kidney Disease Classification Using ML Algorithms

Aya ElMalah [1], Eman A.Shehab [2] Sara A.Shehab [3]
[1] Student, Bioinformatics DP, Faculty of Computer and Artificial Intelligence, University of Sadat City.
[2] Faculty of Computer and Artificial Intelligence, University of Sadat City
[3] Faculty of Computer and Artificial Intelligence, University of Sadat City

Abstract

Chronic kidney failure is one of the most common diseases that threaten the lives of many people and cause death for many. By using artificial intelligence, we predict the disease and classify people into infected and non-infected people. One of the goals is to reduce non-communicable disease-related premature death by a third by 2030. 10-15% of the world's population may have chronic kidney disease (CKD), which is one of the major causes of non-communicable disease morbidity and mortality. In order to reduce the effects of patient health complications like hypertension, anaemia (low blood count), mineral bone disorder, poor nutritional health, acid base abnormalities, and neurological complications with timely intervention through appropriate medications, early and accurate detection of the stages of CKD is thought to be essential. Several studies on the early identification of CKD have been conducted utilising machine learning approaches. They weren't primarily concerned with predicting the exact stages. In this work classification methods are used like support vector classifier, random forest, logistic regression, and decision tree. The results detect that Linear SVC Support Vector Machine achieved high accuracy and Random Forest and Decision tree (100%) and logistic regression achieved (96.8%). A data set with 24 feature and 401 records are used for testing the algorithms. 20% of data set will be used in testing and 80% for training. The proposed work achieves high accuracy when compared with the previous works.

Keywords: *Chronic kidney, Machine Learning, Support Vector Machine, Random Forest.*

1. Introduction

Every 30 minutes[1], your kidneys, which are each about the size of a computer mouse, filter all the blood in your body. They exert a lot of effort to get rid of waste, poisons, and extra fluid. Additionally, they manage blood pressure, promote the generation of red blood cells, maintain bone health, and regulate vital blood molecules. Healthy kidneys are essential for sustaining overall health, yet it is estimated that more than one in seven American people suffer from chronic kidney disease (CKD). Kidney disease, or CKD, impairs the ability of the kidneys to properly filter blood. Due to this, extra fluid and blood waste build up in the body and may result in various health issues like heart disease and stroke. Additional negative effects of CKD on health include:

- Low levels of red blood cells or anemia
- an increase in the frequency of infections
- Low levels of calcium, high levels of potassium, and high levels of phosphorus were found in the blood
- A decrease in eating
- decreased life quality or depression
- Kidney illneses are a leading cause of death in the United States.

About 37 million US adults are estimated to have CKD, and most are undiagnosed. 40% of people with severely reduced kidney function (not on dialysis) are not aware of having CKD. Every 24 hours, 360 people begin dialysis treatment for kidney failure. People with CKD may not have any symptoms or feel unwell. Only particular blood and urine tests can determine for sure if you have CKD. Both the blood creatinine level and the protein content of the urine are measured during these examinations.

In the US, kidney illnesses are one of the main causes of death. Early CKD has neither symptoms nor indicators.

condition's symptoms and prevent further deterioration. Your CKD stage will determine your course of treatment.

The primary therapies are:

- Changes in your lifestyle can help you stay as healthy as possible.
- Medicine can help you manage related issues like high blood pressure and high cholesterol.
- Dialysis is a form of treatment that duplicates some kidney functions and may be required in stage 5 renal disease.
- Advanced CKD patients may also require a kidney transplant.

2. Related Work

Utilizing patient data, various machine learning algorithms have been applied to classify chronic renal disease. On the Indians Chronic Kidney Disease (CKD) dataset, Charleonnan et al. [3] compared the predictive models K-nearest neighbours (KNN), support vector machine (SVM), logistic regression (LR), and decision tree (DT) in order to choose the best classifier for predicting chronic kidney disease. They discovered that SVM has a classification accuracy of 98.3%, which is the highest. On a dataset of 400 observations, Salekin and Stankovic [4] evaluated classifiers like K-NN, RF, and ANN. Five features were chosen for the study's model creation after the implementation of wrapper feature selection. With an RMSE of 0.11 and the greatest classification accuracy of 98%, RF. S. The "Prediction of Chronic Kidney Disease Using Machine Learning Algorithm" study by Tekale et al. [5] used a dataset with 400 occurrences and 14 characteristics. Both decision trees and support vector machines were employed. The dataset has undergone preprocessing, and the 25 features have been decreased to 14. SVM stated as a better model with an accuracy of 96.75%. Using logistic regression, Elastic Net, lasso regression, ridge regression, support vector machines, random forests, neural networks, and

To check for CKD, specific blood and urine tests are required. CKD typically worsens over time.CKD may be treated, and the sooner it is addressed, the better. [2] Kidney failure can develop from CKD. Blood and urine tests can be used to identify chronic kidney disease (CKD).it is frequently only discovered after a regular blood or urine test you get for another issue reveals that your kidneys might not be functioning normally. Chronic kidney disease (CKD) has no known cure; however, treatment can help manage the

In this work, we will work on classifying people into patients with this disease or not by preparing and organizing the dataset and knowing the things that affect the disease and Their presence will be an influential factor in knowing whether the person has the disease or not, then we use machine learning algorithms to predict this disease so we can use artificial intelligence technology to detect chronic kidney disease.

The paper organized as follows: the Related Work is presented in section 2. The proposed algorithm is discussed in section 3. The discussion and experiential results listed in section 4. The paper concluded in section 5.

k-nearest neighbors, Xiao et al. [6] proposed predicting the course of chronic kidney disease and compared the models' performance. They graded the outcome as mild, moderate, or severe using 551 individuals' history data with proteinuria and 18 characteristics. They concluded that, with an AUC of 0.873, logistic regression outperformed other methods. Naive Bayes was used by Priyanka et al. [7] to predict chronic kidney disease. Naive Bayes outperformed other algorithms in their tests, with a 94.6% accuracy rate, including KNN (K-Neares Algorithm), SVM (Support Vector Machines), Decision trees, and artificial neural networks. By examining the data of CKD patients, Yashfi [8] advocated utilising machine learning algorithms to forecast the risk of CKD. Artificial Neural Networks and Random Forest have both been employed. They used RF and ANN to extract 20 of the 25 features. The greatest accuracy for identifying RF was 97.12%. To predict the stages of kidney illness, Rady and Anwar [15] compared the probabilistic neural networks (PNN), multilayer perceptron (MLP), support vector machine (SVM), and radial basis function (RBF) methods. A short dataset and a limited number of features were used in the investigation. The outcome of this study demonstrates that the Probabilistic Neural Networks method has a classification accuracy percentage of

96.7%, which is the highest overall. According to the evaluations mentioned above, several experiments employing machine learning have been done to predict chronic renal disease. The size and timing of data collection are only a few of the many aspects that are crucial to enhancing model performance. This work focuses on the prediction of chronic kidney disease using machine learning models based on a small dataset that was acquired from Kaggle with two classes, ckd and notckd[11-15].

3. Proposed

Preparing the data for classifiers is a crucial step in creating a machine-learning model. Similar missing values exist in the dataset used for this investigation, which must be handled correctly. Additionally, it must be in a modeling-friendly format presented in Figure 1.

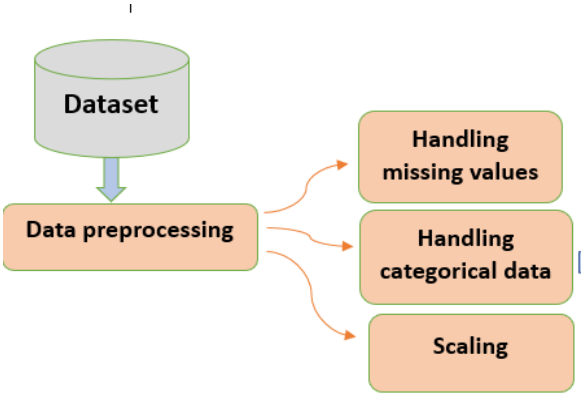

Fig.1. Preparing Dataset

3.1 Data Preprocessing

1) handling missing values Diagnostic test results that would aid in predicting diagnoses or the efficacy of treatments are frequently missing from patient data. The performance of the prediction model is impacted by the missing values. There are several ways for handling missing values, in this study we use mean strategy to handle these missing values
2) Handling Categorical Data: In this step, data has been formatted in accordance with the specifications. the nominal data transformed into 0 and 1 based numerical data. For example, the nominal value of 'Gender' can be labelled as 0-for female and 1-for male. The final CSV file contains all the integer and float values for the various CKD-related features after preprocessing the data.
3) Scaling: Before fitting any models, it is crucial to scale numerical features since certain modelling technique such as support vector machines, logistic regression require scaling. There are different ways for scaling, in this study we use Min-Max scaler.

The process of developing a model utilizing four machine learning algorithms and tenfold cross validation is depicted in Fig.2. Out of the four algorithms, the model with the best performance was chosen as the best machine learning model for each categorization see Fig.2.

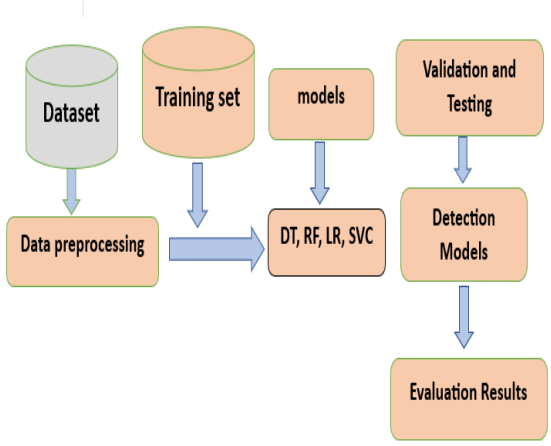

Fig.2.The Proposed Model

In the proposed model, firstly the dataset is preprocessed against error and missing values. Secondly the data after preprocessing is divided into 20% for testing and 80% for training. Finally, selected machine learning algorithm like Decision tree, Random Forest, Linear Regression and Support Vector Machine is applied for classification Stage.

The study's objective was to use machine learning to predict chronic kidney disease. In this research, four machine learning algorithms—Random Forest, Support Vector Machine, Decision Tree and Logistic Regression were employed.

4. Evaluation Results

The crucial process of creating an accurate machine-learning model is performance evaluation. To make sure the model fits the dataset and performs well on unobserved data, the prediction model must be assessed. The purpose of the performance evaluation is to determine how well a model generalizes to unknown or out-of-sample data.The following performance evaluation metrics have been generated in this study: accuracy, precision, recall, f1-score, sensitivity, and specificity. When both the actual value and the anticipated value are positive, this is referred to as a true positive (TP). When both the actual and anticipated values of a data point are negative, the condition is known as a true negative (TN). False

positive (FP) situations occur when the anticipated value is positive even if the actual value of the data point was negative. False negative (FN) situations occur when a data point's real value is positive and its anticipated value is negative. Accuracy: The capacity of the classification algorithm to correctly anticipate the classes of the dataset is implied by the term accuracy. The degree to which the anticipated value resembles the real or theoretical value is measured. Accuracy is typically defined as the ratio of accurate forecasts to total instances. Eq. 3 displays the accuracy equation.[16]

Accuracy=(TP+TN)/TP+FP+TN+FN.

Precision: From the overall predicted values in the actual class, precisely measure the true values that were accurately predicted. Precision measures a classifier's capacity to avoid misclassifying a negative example as positive.

precision equation= TP/TP+FP.

4.1 Logistic Regression Classifier Results

In this result, we observed that Accuracy improved to 96.7% after when achieved 94%. The Confusion Matrix and ROC presented in Figures 3 and 4.

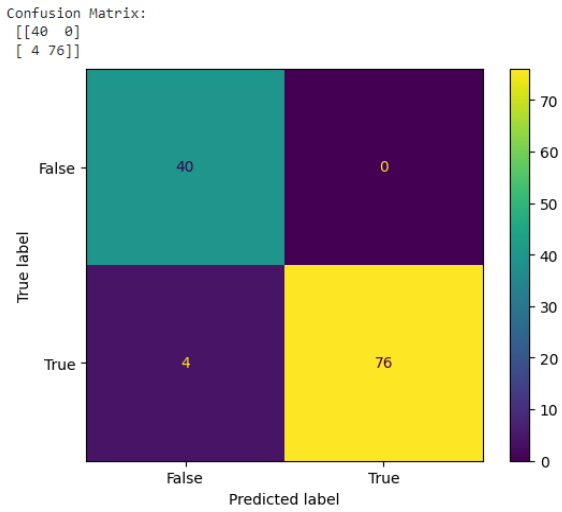

Fig.3. LR Confusion Matrix

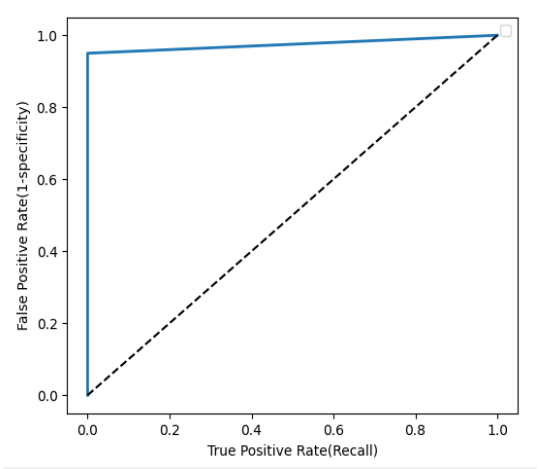

Fig.4. LR ROC

4.2 Decision Tree Classifier Results

In this result, we observed that Accuracy improved to 100% after when achieved 90%. The Confusion Matrix and ROC presented in Figures 5 and 6.

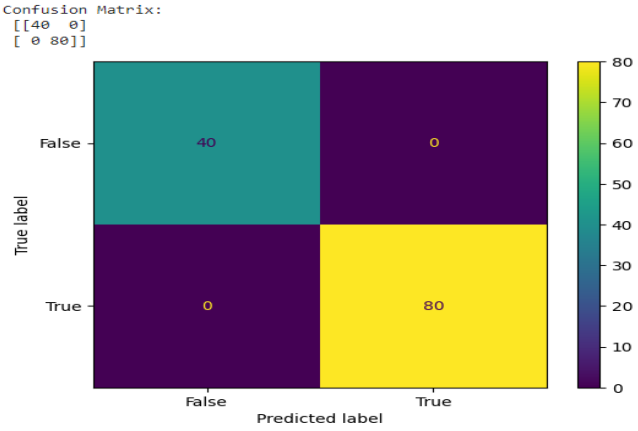

Fig.5. DT Confusion Matrix

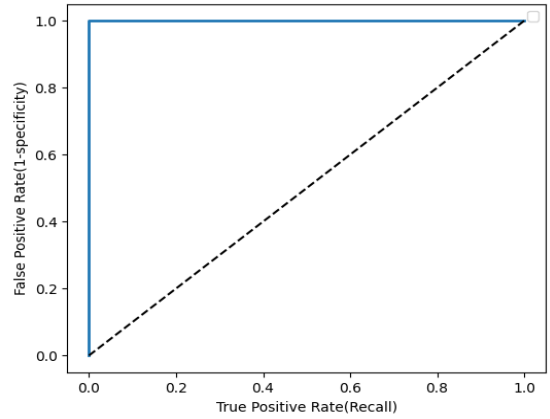

Fig.6. DT ROC

4.3 Random Forest Classifier Results

In this result, we observed that Accuracy 100%. The Confusion Matrix and ROC presented in Figures 7 and 8.

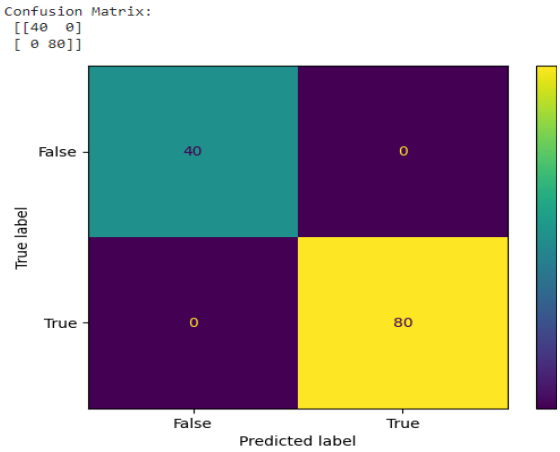

Fig.7. RF Confusion Matrix

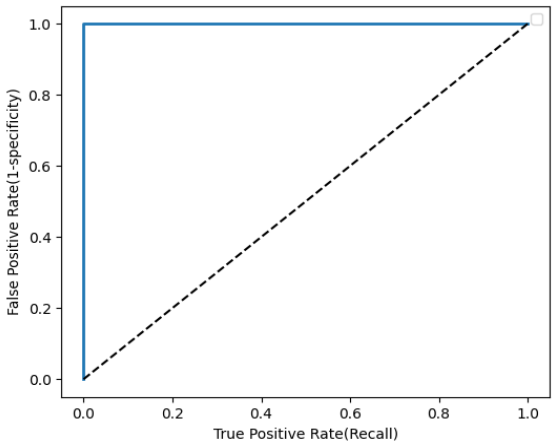

Fig.8. RF ROC

4.4 SVC_ Classifier Results:

In this result, we observed that Accuracy improved to 100% after when achieved 65%. The Confusion Matrix and ROC presented in Figures 9 and 10.

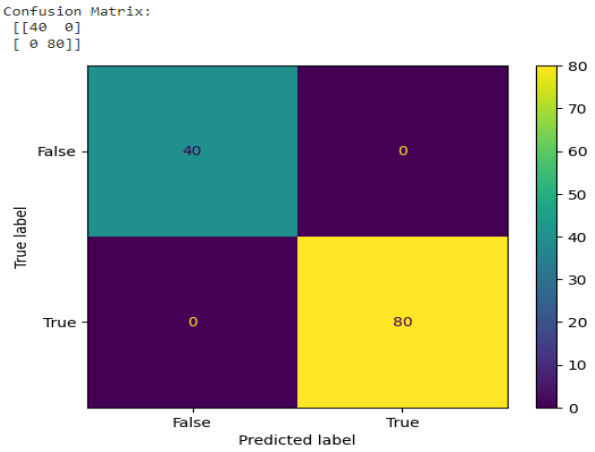

Fig.9. SVC Confusion Matrix

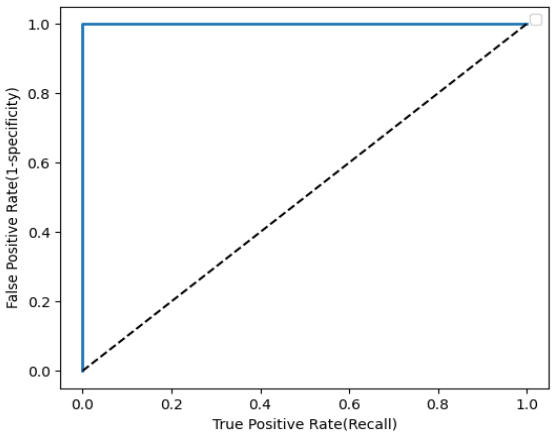

Fig.10. SVC ROC

Table 1 display the performance Matrix for the proposed 4 Machine learning.

Table 1 The Performance Report

Algorithm	F1-Score	Recall	Precision	Accuracy
Logistic Regression	95%	1	91%	96.66%
Decision Tree	1	1	1	100%
Random Forest	1	1	1	100%

SVC	1	1	1	100%

5. Conclusion

For both professionals and patients, making an early forecast is essential to preventing and delaying the progression of chronic renal disease to kidney failure. four machine learning models support vector classifier, random forest, logistic regression and decision tree are used for classification. The binary class's accuracy was 100%. We believe that binary classification work was crucial to understanding the stages of the disease and recommending necessary treatments for the patients to save their lives. Logistic regret generated the lowest result when compared to Random Forest and DT as well.

6. References

[1]Chronic Kidney Disease: Common – Serious – Costly.

[2]Chronic Kidney Disease in the United States, 2021

[3] Charleonnan A, Fufaung T, Niyomwong T,Chokchueypattanakit W, Suwannawach S, Ninchawee N. Predictive analytics for chronic kidney disease using machine learning techniques. Manag Innov Technol Int Conf MITiCON. 2016;80–83:2017.

[4] Salekin A, Stankovic J. Detection of Chronic Kidney Disease and Selecting Important Predictive Attributes. In: Proc. - 2016 IEEE Int. Conf. Healthc. Informatics, ICHI 2016, pp. 262–270, 2016.

[5] Tekale S, Shingavi P, Wandhekar S, Chatorikar A. Prediction of chronic kidney disease using machine learning algorithm. Disease. 2018;7(10):92–6.

[6]Xiao J, et al. Comparison and development of machine learning tools in the prediction of chronic kidney disease progression. J Transl Med. 2019;17(1):1–13

[7]Priyanka K, Science BC. Chronic kidney disease prediction based on naive Bayes technique. 2019. p. 1653–9.

[8]Yashfi SY. Risk Prediction Of Chronic Kidney Disease Using Machine Learning Algorithms. 2020.

[9]Rady EA, Anwar AS. Informatics in Medicine Unlocked Prediction of kidney disease stages using data mining algorithms. Informatics Med. 2019;15(2018):100178

[10]Aqlan F, Markle R, Shamsan A. Data mining for chronic kidney disease prediction. 67th Annu Conf Expo Inst Ind Eng. 2017;2017:1789–94.

[11]Subas A, Alickovic E, Kevric J. Diagnosis of chronic kidney disease by using random forest. IFMBE Proc. 2017;62(1):589–94.

[12]Kapoor S, Verma R, Panda SN. Detecting kidney disease using Naïve bayes and decision tree in machine learning. Int J Innov Technol Explor Eng. 2019;9(1):498–501.

[13]Drall S, Drall GS, Singh S. Chronic kidney disease prediction using machine learning : a new approach bharat Bhushan Naib. Learn. 2014;8(278):278–87.

[14]KadamVinay R, Soujanya KLS, Singh P. Disease prediction by using deep learning based on patient treatment history. Int J Recent Technol Eng. 2019;7(6):745–54.

[15]Ramya S, Radha N. Diagnosis of Chronic Kidney Disease Using. pp. 812–820, 2016.

[16]Acharya A. Comparative Study of Machine Learning Algorithms for Heart Disease Prediction 2017.