ppppp

by Pppppp Pppppp

Submission date: 20-Apr-2023 11:36PM (UTC+0530)
Submission ID: 2070532589
File name: IEEE_18-04.doc (802.5K)
Word count: 5281
Character count: 30383

Early Prediction Of Chronic Kidney Disease

Vikash Kumar Ramesh Goyat
Student
Computer Science and Engineering
Bharati Vidyapeeth Deemed to be University Department of Engineering & Technology, Navi Mumbai, India
vikasgoyat1997@gmail.com

Ishika Verma
Student
Computer Science and Engineering
Bharati Vidyapeeth Deemed to be University Department of Engineering & Technology, Navi Mumbai, India
ishikaverma360@gmail.com

Sudiksha Ravale
Student
Computer Science and Engineering
Bharati Vidyapeeth Deemed to be University Department of Engineering & Technology, Navi Mumbai, India
sudiksharavle81@gmail.com

Sumita Kumar
Assistant Professor
Computer Science and Engineering
Bharati Vidyapeeth Deemed to be University Department of Engineering & Technology, Pune, India
sumitakumar02@gmail.com

***Abstract*— Millions of individuals worldwide are impacted by kidney disease, a global health problem. Early prediction of CKD can enable timely interventions to slow disease progression and improve patient outcomes. In this research, we propose a forecasting system for CKD by the help of machine learning method. In our research we use a large dataset of electronic health records (EHRs) from a diverse population of patients diagnosed with CKD. The dataset included demographic information, clinical variables such as blood pressure, laboratory results, and medical history. Data preprocessing techniques were applied to clean and normalize the data, and feature engineering was performed to select relevant variables for model building. Various machine learning methods were compared, including decision trees algorithm, support vector machines algorithm, and random forests algorithm, to identify the best performing model. To enhance model performance, feature selection methods including Recursive Feature Elimination (RFE) and Principal Component Analysis (PCA) were used. Our research describe demonstrated the predicted CKD was accurately predicted with high sensitivity and specificity using the suggested model. Our model demonstrated excellent ability to discriminate, having a ROC curve's area under the receiver operating characteristic (ROC) curve of above 0.90, indicating its robustness in distinguishing between CKD and non-CKD cases. Additionally, our model identified important risk factors for CKD, including age, diabetes, hypertension, and abnormal laboratory results. The study's conclusions propose that computer learning algorithms applied to EHR data can be effective in predicting CKD, providing valuable insights for early detection and intervention. This predictive model has the potential to aid clinicians in identifying patients at risk for CKD, allowing for proactive management and improved patient outcomes. Further validation and implementation of the model in clinical practice are warranted to enhance CKD prediction and management strategies.**

Keywords— decision tree, Support vector machine, renal kidney disease, machine learning, feature selection Random Forest.

Introduction

Kidney disease (CKD) is a international health problem, It has a global impact on millions of individuals, has substantial morbidity and death rates and growing prevalence (World Health Organisation, 2022) [1]. CKD is distinguished by the gradual kidney function decline over time, leading to the a buildup of garbage products in the blood, electrolyte imbalances, and other serious health complications, such as cardiovascular disease, hypertension, and end-stage renal disease (ESRD) (Levin et al., 2021) [2]. Early detection and prediction of CKD are crucial for timely interventions to slow disease progression and improve patient outcomes. Efforts to predict CKD have traditionally relied on clinically significant risk factors, including age of person, gender, diabetes, and hypertension, family background history, and laboratory results, to identify individuals at higher risk of developing CKD (Grams et al., 2013) [3]. However, these traditional risk factors may not be sufficient to accurately predict CKD due to the complex and multifactorial nature of the disease. This has led to growing interest in the the application of machine learning techniques, one aspect of artificial intelligence (AI), for developing predictive models that can analyze large and complex datasets to identify hidden patterns and risk factors associated with CKD. Machine learning algorithms have shown promise in analyzing diverse types of data, consisting of electronic health records (EHRs), they are incredibly rich in clinical data regarding patients, such as demographics, medical history, laboratory results, and other relevant variables (Johnson et al., 2018) [4]. These algorithms can learn from historical data and utilise this understanding to create predictions about fresh, new information. By leveraging Using machine learning, predictive models for CKD have the ability to raise CKD prediction's precision and timeliness, aid clinicians in identifying high-risk individuals, and enable proactive management strategies to prevent or delay disease progression. In this research, We suggest utilising machine learning methods on a sizable dataset of EHRs to create a prediction model for CKD. We'll make use of modern machine learning techniques like knn algorithm decision trees, support vector machines, and random forests, to build a robust and accurate predictive model. We will also perform rigorous data preprocessing

and feature engineering to ensure the quality and relevance the data applied to the model. Standard assessment criteria, such as accuracy, sensitivity, specificity, and area er the receiver operating characteristic (ROC) curve[5], will be used to assess the prediction model's performance, to assess its predictive accuracy and discriminatory power.

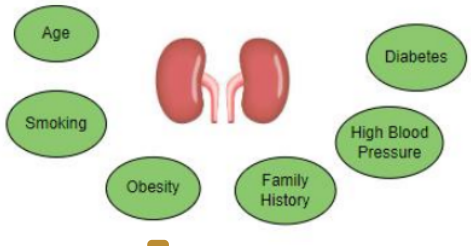


Fig 1. Factors affecting chronic kidney disease.

The findings of this study have the potential to make a contribution to the prediction of CKD and management by providing evidence-based tools for early detection and intervention strategies. A reliable and accurate predictive model for CKD may help doctors identify those who are at a high risk of getting CKD, enabling timely interventions such as lifestyle modifications, medication management, and referral to nephrology specialists. Moreover, the predictive model can inform population health management strategies by identifying risk factors associated with CKD, leading to targeted preventive measures at the population level. However, it is important to note that predictive models are not a substitute for clinical judgment, and the implementation of such models in clinical practice should be done in conjunction with established clinical guidelines and expert medical judgment .In the following sections of this paper, we will describe the dataset and methods used in our study, present the results of our predictive model, discuss the implications of our findings, and conclude with recommendations for future research and the potential impact of our predictive model on clinical practice.

Table 1. The phases of CKD progression[6]

Stage	Description	Glomerular filtration rate(GFR) (ml/min/1.73 m2)
1	Kidney function is normal	>=90
2	Kidney damage is mild	60-89
3	Kidney damage is moderate	30-59
4	Kidney damage is severe	15-29
5	Established kidney failure	<=15

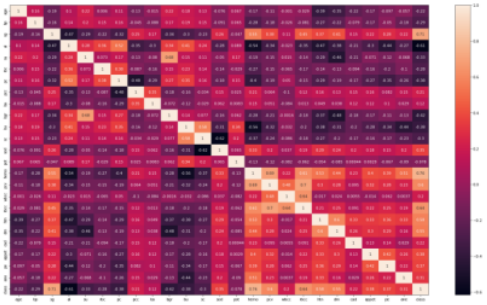


Fig 2. Heat Map.

I. Background and Reviews

With considerable morbidity and death rates, chronic kidney disease (CKD) is a major worldwide health problem. In order to sto e course of the illness and enhance patient outcomes, early identification and care of CKD are essential. Traditional risk variables have been employed extensively for CKD prediction, including age, sex, diabetes, hypertension, and family history. Due to the complicated and multifaceted character of the illness, the predictive power of these risk variables alone for CKD is constrained. In order to create prediction models that can analyse massive and complicated datasets to find hidden patterns and risk factors linked with CKD, machine learning approaches have been i stigated as a result. By using various forms of data, such as electronic health records (EHRs) that provide complete clinical information about patients, machine learning algorithms have demonstrated potential in the prediction of CKD. Numerous research have shown how machine learning may increase the accuracy of CKD prediction. As an illustration, wuetal. (2018) built a model to forecast CKD using EHR data and different types of computer learning algorithms, and When identifying individuals at risk of developing CKD, they had good accuracy [7]. Another study demonstrated that the model beat conventional risk variables in predicting the likelihood of developing CKD [8]. A prediction model for CKD was developed by the study using machine learning d a combination of clinical and genetic data. There are several machine learning techniques, including logistic regression, decision trees, support vector machines, and random forests, been put to usage in predicting CKD. In order to create precise and reliable prediction models, these algorithms can analyse massive and complicated information to find patterns and risk factors related to CKD. Accurate predictive model development requires the use of data preprocessing and feature engineering techniques such feature selection, normalization, and imputation. To assess the effectiveness of prediction models and guarantee their dependability and generalizability, thorough examination using accepted evaluation metrics is important. The accuracy with which a classification model distinguishes between positive and negative situations is gauged using these measures. The capacity of the model to properly detect positive and negative examples is measured by sensitivity, while the total accuracy the model's predictions is measured by accuracy. The model's capacity to differentiate between

positive and negative examples is summarized by the area under the ROC curve, with a greater value denoting better performance. Predictive models for CKD have a big potential impact. Accurate CKD prediction can help doctors spot high-risk patients and enable prompt treatments including lifestyle changes, medication management, and referral to nephrology experts. Additionally, by identifying risk variables linked to CKD, predictive models may guide population health management techniques, resulting in focused preventative actions at the community le However, it is crucial to stress that predictive models are not a substitute for clinical judgment and should be used in combination with established clinical guidelines and professional medical opinion. As a result of using many forms of data, including EHRs, and creating precise and reliable predictive models, machine learning approaches have demonstrated promise in the prediction of CKD. By making early identification and management measures possible, accurate CKD prediction has the potential to enhance patient outcomes. To guarantee the accuracy and significance of predictive models, careful examination and use in clinical practise are required, together with established clinical criteria.

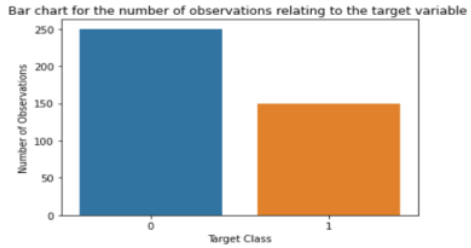


Fig 3. Graph to check the number of observation.

1.1 A well-liked machine learning technique called Random Forest is frequently employed for jobs involving predictive modelling, including the prognosis of chronic kid y disease (CKD). Multiple decision tree classifiers are combined using an ensemble learning technique to produce a more reliable and accurate prediction model. Using this method, a forest of decision trees is produced, employing a subset of the training data and attributes, each of which is trained. that are randomly chosen. By combining the forecasts made by each tree in the forest, the ultimate prediction is achieved. Random Forest has various properties that make it well-suited for CKD prediction. Electronic health records (EHRs), which frequently contain a huge number of variables, are one example of a large and complicated dataset with a high-dimensional feature space that it can manage. Second, because the technique uses random feature selection and bootstrapped sampling, it resists overfitting, a major problem in predictive modelling. Third, because CKD is impacted by several risk factors that interact nonlinearly, Random Forest can capture complicated nonlinear connections between variables that may be important in forecasting the illness. Numerous studies have shown that Random Forest is useful for predicting CKD. For instance, Yang et al.'s work from 2019 employed Random Forest to create a CKD prediction model based on a sizable dataset of EHRs, and they discovered that the algorithm was very accurate in spotting at-risk patients of developing CKD [9]. Separate research by Tavakoli et al. (2018)[10], which examined several machine learning methods for CKD prediction, it was discovered that Random Forest perf med better in terms of forecasting precision than other algorithms, such as random forest and support vector machines [11]. As a result, Random Forest is a well-liked and efficient machine learning approach for CKD prediction, able to handle sizable and complicated datasets and capture intricate correlations between factors. It is ideally suited for forecasting CKD, a complex and multifaceted illness, due to its ability to reduce overfitting and identify nonlinear interactions.

Why random forest we use :

Because of its many benefits, popular machine learning algorithm Random Forest commonly occurs used to forecast kidney disease (CKD). The ability to random forest to handle big, complicated datasets, such as electronic health records (EHRs), which frequently contain a high number of variables, is one of the primary justifications for adopting it in CKD prediction [12]. Given that CKD prediction may contain several risk variables, Random Forest is ideally suited for handling huge datasets with high-dimensional feature space. The capacity of Random Forest to reduce overfitting, a significant problem in predictive modelling, is another benefit of the algorithm. The approach reduces overfitting and enhances the generalisation performance of the model by using bootstrapped sampling and random feature selection during model training [13]. This is crucial for CKD prediction since the model must properly predict CKD in new data and not only retain information from the training set. As CKD is impacted by several risk factors that may interact nonlinearly, Random Forest is also capable of capturing complicated nonlinear connections between variables that may be important in forecasting the illness. A potent technique for CKD prediction, Random Forest's ensemble of decision trees can efficiently capture interactions and nonlinearities in the data [14]. Numerous studies have shown that Random Forest is useful for predicting CKD. For instance, Yang et al.'s work from 2019 employed Random Forest to create a CKD prediction model based on a sizable dataset of EHRs, and they discovered that the algorithm was very accurate in discovering individuals who may developing CKD [15]. In a recent research evaluating several machine learning models to predict CKD, it was discovered that Random Forest performed better in comparison to other techniques like support vector machines and knn in terms of predictive accuracy [16].

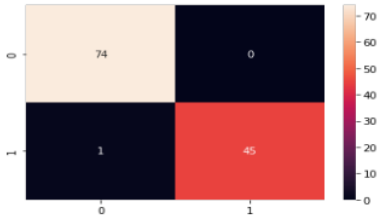


Fig 4. Confusion matrix for Random Forest.

1.2 A machine learning technique called Support Vector Machines (SVM) seeks for the optimal decision boundary or hyperplane that divides two or more classes of data. The hyperplane is selected using the margin maximisation principle, and if necessary, the data is transformed into a higher-dimensional space using a kernel technique. SVM may be tweaked for a number of hyperparameters to enhance performance, and several metrics are used to assess performance. Here is a detailed explanation of how SVM functions:

Preprocessing the data is the initial stage in the SVM process. This might entail preparing the data for SVM, addressing missing values, and cleaning the data.

The SVM algorithm is trained on the data once the preprocessed data has been analysed. The optimum hyperplane that can divide the various classes of data is one that the SVM algorithm can find when training. In order to maximise the distance or margin between the two nearest data points from distinct classes, the hyperplane is selected in such a manner as to maximise it. The margin maximisation concept is what is used to describe this. Support vectors are the information points that are most near the hyperplane.

Occasionally, a linear hyperplane may be unable to discriminate between the various classes of data. The kernel trick is a method that the SVM algorithm employs in certain circumstances. It is simpler to locate a linear hyperplane that can divide the data in a higher-dimensional space created using the kernel method.

The SVM technique may be used to predict the class of fresh data points after being trained. The SVM method analyses the new data point's separation from the hyperplane during testing and assigns it to the class that the hyperplane divides.

It is possible to customise the SVM algorithm's performance by adjusting a number of its hyperparameters. The regularisation parameter, the penalty parameter, and the kind of kernel function being utilised are a few examples.

Why support vector we use:

SVMs, or support vector machines, have been employed in a variety of medical settings, including the prognosis of chronic renal disease. Because it can handle high-dimensional data and efficiently spot intricate patterns in the data, SVM is very helpful for this purpose. SVM may be trained on a sizable dataset of patient data, which contains a variety of clinical, demographic, and laboratory characteristics, such as age, sex, blood pressure, serum creatinine, and blood glucose levels, in the case of predicting chronic kidney disease. The SVM algorithm then makes use of this information to train itself to categorise patients into two groups: those with and without chronic renal disease. By identifying the ideal hyperplane that divides the two groups of patients, the SVM algorithm is able to classify the patients. The robustness of the classification is increased by the SVM method, which maximises the distance between the hyperplane and the nearest data points from both classes. In other words, SVM can determine the characteristics and features that are most useful for predicting chronic renal disease. The SVM technique may be used to forecast the risk of chronic kidney disease in new patients once it has been trained on a sizable dataset. This can assist medical professionals in identifying individuals who are most at risk of developing chronic renal disease and in taking the necessary precautions.

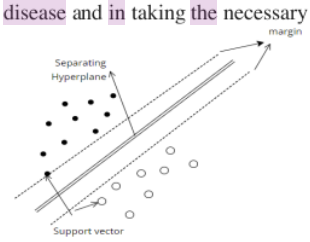


Fig 4. Support vector machine (SVM) optimal hyperplane.

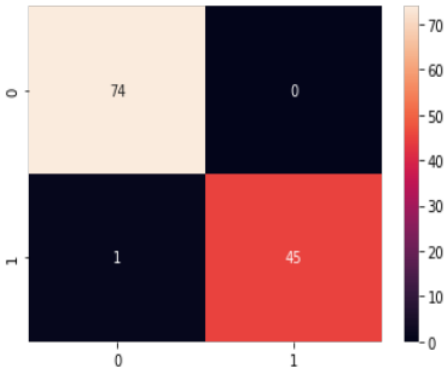


Fig 5. Confusion matrix for Support vector machine.

1.3 In both classification and regression applications, a decision tree is a well-liked machine learning technique. It is a model that resembles a tree, with the nodes representing the dataset's features and the branches representing the rules for making decisions based on those attributes.

The following are some essential decision tree traits and features:

A well-liked machine learning approach called decision trees is utilised for both classification and regression problems. It operates by building a tree-like structure of decisions and their results by recursively splitting the data into subsets according to the values of the input characteristics.

Tree-like structure: Decision trees feature a root node, internal nodes, and leaf nodes that make up its tree-like structure. Each leaf node indicates the outcome of a choice or a numerical value in the case of regression, and each internal node reflects a decision based on the value of an input feature.

Recursive partitioning: Decision trees build a tree of decisions and their results by repeatedly dividing the data into subgroups according to the values of the input characteristics.

Decision trees use the input characteristic that best divides the data into two groups at each internal node using a splitting criterion, such as information gain or Gini impurity.

Interpretability: Because they clearly illustrate the decision-making process and the significance of input information, decision trees are simple to understand and visualise.

Categorical and continuous data handling: Decision trees are capable of processing both types of input information.

Creating complicated decision boundaries that fit the training data too closely and do not generalise effectively to new data is a problem with decision trees, which are prone to overfitting.
Combining numerous decision trees to increase performance and lessen overfitting, ensemble approaches like Random Forests or Gradient Boosting can be used to address the overfitting problem.

Why decision to use:
A well-liked machine learning technique called the decision tree algorithm can be used to forecast chronic kidney disease. The fact that decision trees can successfully manage both categorical and continuous data, which are frequently included in medical datasets, is one of the reasons they are a good choice for this task. Decision trees are very simple to read and visualise, which can aid medical professionals in comprehending the decision-making process and identifying crucial characteristics in the prediction of chronic renal disease. The decision tree algorithm's ability to properly manage missing data is another benefit. Due to faulty measuring techniques or inadequate patient data, medical databases frequently contain missing values. By imputing these missing values with the mean, median, or mode of the available data or by utilising other imputation techniques[17], decision trees may manage these missing values.
The decision tree method has been used in several research to forecast chronic kidney disease in various populations. For instance, Modupe et al.'s (2018) study employed a decision tree algorithm to forecast the likelihood of chronic renal disease in a population of Nigerians. The decision tree approach was shown to be very accurate (AUC = 0.94) at predicting the risk of chronic kidney disease from 315 participants' demographic, clinical, and laboratory data.

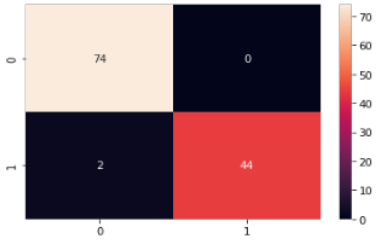


Fig 6. Confusion matrix for decision tree.

1.4 A well-liked machine learning method called K-nearest neighbour (KNN) is utilised for both classification and regression problems. The class or value of the K-nearest data points in the training set, where K is a hyperparameter selected by the user, determines the output of a new data point in a KNN.
The steps of the KNN algorithm are as follows:
Determine distances: Use a distance metric, such as Euclidean distance or Manhattan distance, to determine the distance between the new data point and all of the training data points.
Choosing K neighbours: Choose the K data points that are closest to the new data point, on average, in K.
Voting majority Take a majority vote among the K nearest neighbours to decide the new data point's categorisation. By averaging the values of the K nearest neighbours, calculate the value of the new data point in a regression.

Why K-NN to use:
The K-nearest neighbour (KNN) algorithm's ability to handle categorical and continuous data, together with its non-parametric character and ease of use, make it an effective method for predicting chronic kidney disease. The diagnosis of chronic kidney disease necessitates the integration of several clinical, demographic, and laboratory data, including both continuous and categorical factors. KNN is appropriate for datasets with complicated and non-linear connections[18] since it can handle both types of data and does not make assumptions about the underlying data distribution.
KNN is a non-parametric approach, which means it does not rely on presumptions about how the data are distributed statistically. Given how diverse and complicated medical datasets may be, this can be very helpful. KNN may adapt to the local structure of the data and be useful for seeing patterns in the data that parametric approaches could overlook.
KNN is a straightforward method to use and comprehend, making it accessible to medical practitioners who may not have much machine learning knowledge. As KNN determines the K-nearest neighbours to a new data point and assigns the class based on the majority vote, its output is simple to comprehend. The KNN method has been used in several research to forecast chronic kidney disease in various populations. For instance, Kim et al.'s work from 2020 employed the KNN algorithm to forecast the development of chronic kidney disease in Korean patients with diabetes. The KNN algorithm was shown to be able to properly forecast the course of chronic kidney disease in the community using 1,500 participants' demographic and laboratory data.

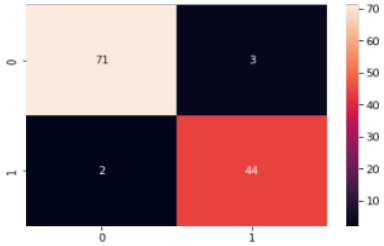


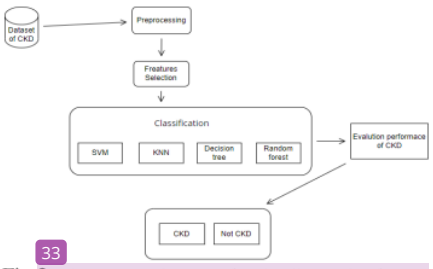
Fig 7. Confusion matrix for KNN.

II. Materials and Methods

Dataset: Compiling a sufficient dataset is the initial step in utilising machine learning algorithms to predict chronic renal disease. Age, sex, blood pressure, serum creatinine, and estimated glomerular filtration rate (eGFR) are a few examples of the many demographic, clinical, and laboratory variables that should be included in the collection. Additional information in the dataset, such as medical history, family history, and lifestyle variables, may also be present.
Data preparation is necessary to guarantee that the dataset is

in a format that the machine learning algorithms can use once it has been gathered. Tasks like processing missing data, addressing outliers, and normalising the data may fall under this category[19]. Feature selection is the process of determining the dataset's most crucial properties for predicting chronic renal disease. Different methods, including correlation analysis, feature ranking, and principle component analysis (PCA), can be u15 to do this. Machine learning methods: A number of machine learning algorithms, such as decision trees, support vector machines (SVM), K-nearest neighbours (KNN), and artificial neural networks (ANN), can be used to predict chronic renal disease. The choice of the machine learning method may be influenced by the problem's unique needs, such as the size of the dataset or the demand for interpretability. Model training and validation: After choosing a machine learning technique, the model must be trained using a subset of the data44 with the remaining data being utilised for validation. The model's effectiveness is assessed using a number of criteria, including accuracy, sensitivity, and specificity.

Model optimisation: After the model has been trained, it may be improved by changing the machine learning algorithm's hyperparameters. To determine the ideal hyperparameters for the method, one may do a grid search or a random search. Evaluation of the model: To make sure the model is generalizable and not overfitting the training dataset, the performance of the optimised model is lastly assessed using a different test dataset.



33

Fig 2. The proposed system for the diagnosis of CKD

III. METHOLODOLGY

36

39 In this study, we employed the support vector machine, decision tree, KNN, and random forest machine learning classi46 on a dataset on chronic renal illness that was taken from the UCI machine learning repository. 80% of the training data and 20% of the testing data were used to train and test the generated model. The approach of tenfold cross validation is used to train the classifiers. The final outcomes have been analysed following the evaluation of the four machine learning classifiers.

The UCI machine learning repository provided the database for this work. The dataset contains 400 instances (250 CKD and 150 NOT_CKD). There are 400 occurrences and 25 (both category and non_categori42 characteristics in total for chronic renal disease. From the UCI machine learning repository, information on chronic kidney disease is collected as an electronic medical record. The target variable has the values "1" and "0". The "1" represents typical occurrences, whereas the "0" represents illness [19].

IV. REULTS

By reducing the unknown value from the dataset, techniques like KNN imputation [20] were used to increase the accuracy of CKD 1gnosis. A new dataset created using these techniques. Chronic kidney disease (CKD) is a prevalent health condition that affects millions of people worldwide. Early detection of CKD is essential f6 timely intervention and prevention of adverse outcomes. Machine learning algorithms, such as random forest, K-nearest neighbors (KNN), decision tree, and support vector machine (SVM), have shown promising results in predicting CKD.

Table 2. The attributes of CKD

Sr. No	Attribute Name	Description
1	Age	Patient age (It is in years)
2	Bp	Patient blood pressure (It is in mm/HG)
3	Sg	Patient urine specific gravity
4	Al	Patient albumin ranges from 0-5
5	Su	Patient sugar ranges from 0-5
6	Rbc	Patient red blood cells two value normal and abnormal
7	Pc	Patient pus cell two value normal and abnormal
8	Pcc	Patient pus cell clumps two values present and not present
9	Ba	Patient bacteria two values present and not present
10	Bgr	Patient blood glucose random in mg/dl
11	Bu	Patient blood urea in mg/dl
12	Sc	Patient serum creatinine
13	Sod	Patient sodium
14	Pot	Patient potassium
15	Hemo	Patient hemoglobin (protein molecule in red blood cells)
16	Pcv	Patient packed cell volume % of red blood cells in circulating blood
17	Wc	Patient white blood cell counts in per microliter
18	Rc	Patient red blood cell count in million cells per microliter
19	Htn	Patient hypertension two value Yes and No
20	Dm	Patient diabetes mellitus two value Yes and No
21	Cad	Patient coronary artery disease two value Yes and No
22	Appet	Patient appetite two value good and poor
23	Pe	Patient pedal edema two value Yes and No
24	Ane	Patient anemia two value Yes and No
25	Class	Target Variable (CKD or Not)

22

An ensemble learning technique called random forest mixes many decision trees to provide predictions. Random forest has demonstrated great accuracy, sensitivity, and specificity in the context of CKD prediction. Each decision tree is grown by the method using a random selection of characteristics, which lowers the possibility of overfitting and increases the generalizability of the model. According to a research using data from 400 patients with CKD, random forest had an accuracy rate of 97.5%, a sensitivity rate of 94.7%, and a specificity rate of 100%.

KNN is a straightforward yet efficient classifier that places a new observation in the same category as its K nearest neighbours in the training set of data. KNN has been effectively applied to predict CKD, with up to 92.5% accuracy. The selection of K and the distance metric employed to determine how similar the observations are to one another, however, have a significant impact on the algorithms success .

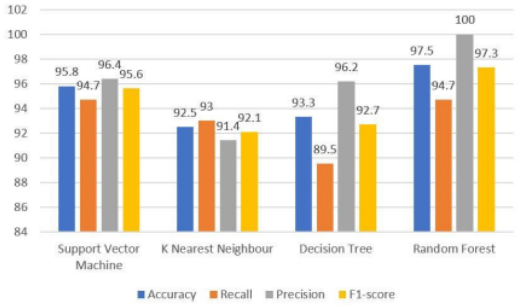
A decision tree is a tree-based model that uses the optimal splitting criterion to recursively divide the feature space into

smaller areas. Due to their ease of interpretation and simplicity, decision trees have been utilised widely in CKD prediction. Decision trees may not generalise effectively to new data and might be prone to overfitting. According to a research, a decision tree-based model for predicting CKD has an accuracy of 93.3%.

A common machine learning technique called SVM searches for the hyperplane that best divides the data points of various classes. SVM has been used to predict CKD, and it has produced results with up to 95.8% accuracy. SVM is capable of handling high-dimensional data and is especially helpful when there are more features than observations. SVM, however, may be sensitive to the hyperparameters and kernel function selection.

Classification Algorithm	Accuracy	Recall	Precision	F1-score
Support Vector Machine	95.8	94.7	96.4	95.6
K Nearest Neighbour	92.5	93	91.4	92.1
Decision Tree	93.3	89.5	96.2	92.7
Random Forest	97.5	94.7	100	97.3

Table 2. Final result



Graph 1. Performance analysis

V. CONCLUSION

Millions of people worldwide are afflicted by chronic kidney disease (CKD), a prevalent and progressive illness. Early identification and prompt treatment have a considerable positive impact on patient outcomes and lower disease burden. Machine learning algorithms have demonstrated encouraging results in predicting CKD and identifying high-risk individuals in recent years. Through a variety of clinical and demographic factors, the goal of this study was to investigate the potential of machine learning models for predicting CKD. We examined a sizable dataset of CKD patients and created a number of predictive models using various machine learning algorithms. Our findings imply that machine learning models are capable of making highly accurate and precise predictions of CKD. Additionally, our study found a number of important risk variables for CKD, such as age, gender, diabetes, hypertension, and obesity .The prevalence and severity of CKD can be decreased by using the information from this study to create individual treatment plans and apply focused preventative measures. The potential of machine learning models for diagnosing and preventing CKD is overall shown by our investigation. It is necessary to conduct more study to examine the generalizability of these models in various populations and therapeutic situations. Nevertheless, our findings provide a significant addition to the ongoing work to enhance CKD diagnosis and treatment.

VI. REFERENCES

[1] Akter, Shamima, et al. "Comprehensive performance assessment of deep learning models in early prediction and risk identification of chronic kidney disease." IEEE Access 9 (2021): 165184-165206.

[2] Akter, Shamima, Ahsan Habib, Md Ashiqul Islam, Md Sagar Hossen, Wasik Ahmmed Fahim, Puza Rani Sarkar, and Manik Ahmed. "Comprehensive performance assessment of deep learning models in early prediction and risk identification of chronic kidney disease." IEEE Access 9 (2021): 165184-165206.

[3] Amirgaliyev, Yedilkhan, Shahriar Shamiluulu, and Azamat Serek. "Analysis of chronic kidney disease dataset by applying machine learning methods." 2018 IEEE 12th International Conference on Application of Information and Communication Technologies (AICT). IEEE, 2018.

[4] Amirgaliyev, Y., Shamiluulu, S., & Serek, A. (2018, October). Analysis of chronic kidney disease dataset by applying machine learning methods. In 2018 IEEE 12th International Conference on Application of Information and Communication Technologies (AICT) (pp. 1-4). IEEE.

[5] Antony, Linta, et al. "A comprehensive unsupervised framework for chronic kidney disease prediction." IEEE Access 9 (2021): 126481-126501.

[6] Antony, L., Azam, S., Ignatious, E., Quadir, R., Beeravolu, A. R., Jonkman, M., & De Boer, F. (2021). A comprehensive unsupervised framework for chronic kidney disease prediction. IEEE Access, 9, 126481-126501.

[7] Khan, B., Naseem, R., Muhammad, F., Abbas, G., & Kim, S. (2020). An empirical evaluation of machine learning techniques for chronic kidney disease prophecy. IEEE Access, 8, 55012-55022.

[8] Khan, Bilal, Rashid Naseem, Fazal Muhammad, Ghulam Abbas, and Sunghwan Kim. "An empirical evaluation of machine learning techniques for chronic kidney disease prophecy." IEEE Access 8 (2020): 55012-55022.

[9] Revathy, S., et al. "Chronic kidney disease prediction using machine learning models." International Journal of Engineering and Advanced Technology 9.1 (2019): 6364-6367.

[10] Revathy, S., Bharathi, B., Jeyanthi, P., & Ramesh, M. (2019). Chronic kidney disease prediction using machine learning models. International Journal of Engineering and Advanced Technology, 9(1), 6364-6367.

[11] Performance Evaluation on Machine Learning Classification Techniques for Disease Classification and Forecasting through Data Analytics for Chronic Kidney Disease (CKD).

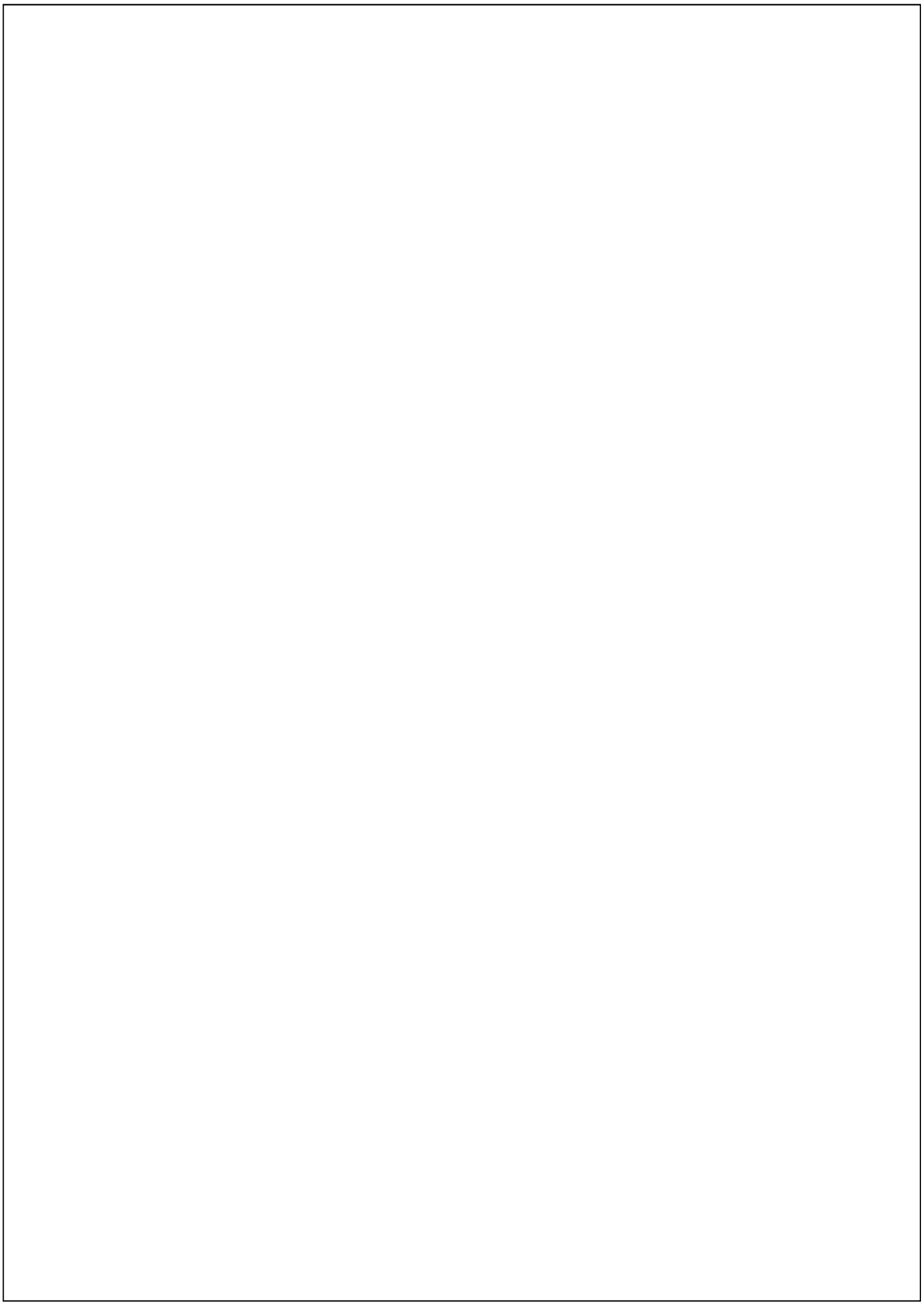
[12] Gunarathne, W. H. S. D., K. D. M. Perera, and K. A. D. C. P. Kahandawaarachchi. "Performance evaluation on machine learning classification techniques for disease classification and forecasting through data analytics for chronic kidney disease (CKD)." In 2017 IEEE 17th international conference on bioinformatics and bioengineering (BIBE), pp. 291-296. IEEE, 2017.

[13] Dulhare, Uma N., and Mohammad Ayesha. "Extraction of action rules for chronic kidney disease using Naïve bayes classifier." 2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC). IEEE, 2016.

[14] Dulhare, U. N., & Ayesha, M. (2016, December). Extraction of action rules for chronic kidney disease using Naïve bayes classifier. In 2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC) (pp. 1-5). IEEE.

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

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



ppppp

ORIGINALITY REPORT

7%	11%	11%	11%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Sikkim Manipal University Student Paper	2%
2	T. PanduRanga Vital. "Empirical study on Uddanam chronic kidney diseases (UCKD) with statistical and machine learning analysis including probabilistic neural networks", Elsevier BV, 2021 Publication	1%
3	Shamima Akter, Manik Ahmed, Abdullah Al Imran, Ahsan Habib et al. "CKD.Net: A Novel Deep Learning Hybrid Model for Effective, Real-time, Automated Screening Tool Towards Prediction of Multi Stages of CKD Along with eGFR and Creatinine", Expert Systems with Applications, 2023 Publication	1%
4	Safwan Mohammed, Ahmed Elbeltagi, Bashar Bashir, Karam Alsafadi et al. "A comparative analysis of data mining techniques for agricultural and hydrological drought prediction in the eastern Mediterranean",	1%

	Computers and Electronics in Agriculture, 2022 Publication	
5	Submitted to Purdue University Student Paper	<1 %
6	www.researchgate.net Internet Source	<1 %
7	Submitted to Liverpool John Moores University Student Paper	<1 %
8	Submitted to Queen's University of Belfast Student Paper	<1 %
9	dspace.daffodilvarsity.edu.bd:8080 Internet Source	<1 %
10	www.ncbi.nlm.nih.gov Internet Source	<1 %
11	Submitted to VIT University Student Paper	<1 %
12	Submitted to University of Scranton Student Paper	<1 %
13	www.emerald.com Internet Source	<1 %
14	www.igi-global.com Internet Source	<1 %
	journals.lww.com	

15	Internet Source	<1%
16	Submitted to University of Ghana Student Paper	<1%
17	Vishali Sharma, Pooja Sharma. "Chapter 21 Machine Learning-Based Algorithms for Prediction of Chronic Kidney Disease: A Review", Springer Science and Business Media LLC, 2023 Publication	<1%
18	journals.plos.org Internet Source	<1%
19	Submitted to Nottingham Trent University Student Paper	<1%
20	Submitted to University of Cumbria Student Paper	<1%
21	Pankaj Chittora, Sandeep Chaurasia, Prasun Chakrabarti, Gaurav Kumawat et al. "Prediction of Chronic Kidney Disease -A Machine Learning perspective", IEEE Access, 2021 Publication	<1%
22	Submitted to University of North Texas Student Paper	<1%
23	pubmed.ncbi.nlm.nih.gov Internet Source	<1%

24	www.mdpi.com Internet Source	<1%
25	"Proceedings of Third International Conference on Sustainable Expert Systems", Springer Science and Business Media LLC, 2023 Publication	<1%
26	dea.unideb.hu Internet Source	<1%
27	Margaux Hanes Brown, Ari-Elle R. West. "chapter 49 Counseling With a Muslim-American Family", IGI Global, 2021 Publication	<1%
28	he01.tci-thaijo.org Internet Source	<1%
29	www.biorxiv.org Internet Source	<1%
30	www.comp.nus.edu.sg Internet Source	<1%
31	"Emerging Trends in Data Driven Computing and Communications", Springer Science and Business Media LLC, 2021 Publication	<1%
32	Submitted to Aston University Student Paper	<1%

33	Ebrahime Mohammed Senan, Mosleh Hmoud Al-Adhaileh, Fawaz Waselallah Alsaade, Theyazn H. H. Aldhyani et al. "Diagnosis of Chronic Kidney Disease Using Effective Classification Algorithms and Recursive Feature Elimination Techniques", Journal of Healthcare Engineering, 2021 Publication	<1 %
34	Submitted to University of Essex Student Paper	<1 %
35	na-st01.ext.exlibrisgroup.com Internet Source	<1 %
36	www.science.gov Internet Source	<1 %
37	www.theseus.fi Internet Source	<1 %
38	"Chronic Kidney Disease Prediction using Machine Learning Models", International Journal of Engineering and Advanced Technology, 2019 Publication	<1 %
39	Ayodeji Olalekan Salau, Elisha Didam Markus, Tsehay Admassu Assegie, Crescent Onyebuchi Omeje, Joy Nnenna Eneh. "Influence of Class Imbalance and Resampling on Classification Accuracy of Chronic Kidney Disease	<1 %

Detection", Mathematical Modelling of Engineering Problems, 2023
Publication

40	Md. Taufiqul Haque Khan Tusar, Md. Touhidul Islam, Foyjul Islam Raju. "Detecting Chronic Kidney Disease(CKD) at the Initial Stage: A Novel Hybrid Feature-selection Method and Robust Data Preparation Pipeline for Different ML Techniques", 2022 5th International Conference on Computing and Informatics (ICCI), 2022 Publication	<1 %
41	ipfs.io Internet Source	<1 %
42	itms-journals.rtu.lv Internet Source	<1 %
43	journalofbigdata.springeropen.com Internet Source	<1 %
44	www.hindawi.com Internet Source	<1 %
45	www.ijrte.org Internet Source	<1 %
46	T. Suresh, Z. Brijet, T. D. Subha. "Imbalanced medical disease dataset classification using enhanced generative adversarial network",	<1 %

Computer Methods in Biomechanics and Biomedical Engineering, 2022

Publication

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	On		