Chronic Kidney Disease Detection Using RFA

Prof. Sachin Wakurdekar[1], Munish Goswami[2], Anupama Sharma[3], Harshit Gupta[4]
[1,2,3,4]*Department of computer engineering, Bharati Vidyapeeth (Deemed to be university), College of Engineering, Pune*

Abstract **- Timely Diagnosis in healthcare is very crucial and important. According to the World Health Organization (WHO) at least 8.8 million people die of cancer annually due to late diagnosis. In the present times and Technology, we can predict a diagnosis given the Important Features (example, red blood cell count, white blood cell count, etc). We can predict the diagnosis only after Analysis of Large Data Set of people having and not having a Disease. Currently, kidney disease is a major problem. Because there are so many people with this disease. The objective is to develop a simple yet powerful tool, Chronic Kidney Disease Predictor, which predicts whether a Person has CKD or not (through a Data Set of 400 people) and provides a prediction which can be confirmed by performing a KFT, Medical test. With the help of Machine Learning and use of several Algorithms like Random Forest, SVM etc we can be able to Predict whether a Patient has C.K.D. or not. Users can be a doctor or any person who is able to read Medical Reports and can send the result to the Patient in just one click. It is especially very useful for people in health care because of time diagnosis.**

Index Terms **- Machine Learning, WHO- World Health Organization, CKD- Chronic Kidney Disease, SVM- Support Vector Machine, KFT- Kidney Function Test, LR- Logistic Regression, KNN- K Nearest Neighbours.**

I.INTRODUCTION

CKD is the non-communicable disease (NCDs) that has significantly contributed to increased morbidity, mortality, and an admission rate of patients throughout the world. Chronic kidney disease is a condition in which the kidney structure is unusual, or function reduces for 3 months and more with a reduced glomerular filtration rate. Nowadays, (CKD) is becoming a challenging public health problem with increasing prevalence worldwide, highly burdening low-income countries, where detection, prevention and treatment rates remain low. According to the global burden disease project, CKD has now become a quickly expanding and killing disease all over the world. A report from 1990 to 2013 indicated the globally yearly life loss of CKD increased by 90% and it is known as the 13th leading death cause in the world. Currently, 850 million people throughout the world are likely to have kidney diseases from different factors and the report world kidney day of 2019 indicated that the disease become a serious disease from which at least 2.4 million dies per year and now it is the 6th fastest-growing cause of death worldwide.

II.MOTIVATION

The report from different health organizations and different statistical analysis indicates that the number of people suffering from CKD is increasing from time to time worldwide. Especially in developing countries like Ethiopia, chronic kidney disease is becoming a serious disease. Many researchers have conducted different research on how to manage the progress of chronic kidney disease locally and globally. Most globally conducted research on chronic kidney disease prediction focuses on the identification of the presence or absence of the disease but not on the identification of the severity of the diseases. Therefore, we were inspired to work on the CKD prediction using machine learning techniques because machine learning can analyze medical data and predict the severity of the disease correctly by simply inserting common clinical predictors. Using machine-learning techniques assists and enables the experts to make fast decisions and take necessary action to treat patients and give an appropriate recommendation based on the severity of the disease. Collecting the new dataset from locally collected patient history data, the study will contribute to the building of a better version of CKD prediction to save the time and cost of treatment with the help of machine learning predictive models.

III.OBJECTIVE AND SCOPE

Timely Diagnosis is a very valuable thing. It can save many lives. New technologies are introduced every day which can help us have a timely diagnosis. Hence this Chronic Kidney Disease Predictor Tool using Machine Learning & Python is introduced. Machine Learning is a vast field concerned with the study of enormous and several variable data and grown from the study of pattern recognition and computational learning theory in artificial intelligence, having several computational methods, algorithms and techniques for interpretation and prediction. In Medical Science's viewpoint, Machine Learning techniques have shown success in prediction and diagnosis of numerous critical diseases. In this strategy some set of features are used for the representation of every instance in any dataset that is used. Furthermore, human professionals and experts are limited in finding hidden patterns from data. Hence, the alternative is to use computational methods to investigate the fresh data and mine exciting information for the decision-maker. This project is aimed to develop a simple yet powerful tool, chronic kidney disease, which predicts whether a Person has CKD or not (through Data of 400 people is Analysed) and provides a prediction which can be confirmed by performing a KFT, Medical test. With the help of these tools, a doctor can be able to get the diagnosis in time with the help of prediction that is extracted in a very way, by just entering the values and clicking a button and the doctor even has the option to send the prediction directly to the Patient via email. The objective is to develop a simple yet powerful tool, Chronic Kidney Disease Predictor, which predicts whether a Person has CKD or not (through a Data Set of 400 people) and provides a prediction which can be confirmed by performing a KFT, Medical test. With the help of Machine Learning and use of several Algorithms like Random Forest, SVM, KNN etc. We can be able to Predict whether a Patient has C.K.D or not. The scope of this project is to implement a reliable Tool which gives us a very accurate prediction that can be used by a Doctor or any person who is able to read Medical Reports and instantly share it with the Patient and help the patients in the coming future.

IV.LITERATURE SURVEY

This section discusses the review of necessarily related research papers to chronic kidney disease prediction to clearly understand general techniques, methods, and results of current studies. Many researchers have been researching the system that can predict the diseases and still the health sectors are left with huge researchable areas that need the intervention of machine learning suitable for the classification and prediction of disease such as machine learning, deep learning. Different related works have been conducted locally and globally by different researchers on chronic kidney disease prediction using machine-learning techniques.

Review of Existing Models, Approaches, Problems

- A. Charleonnan et al. presented predictive models such as K-nearest neighbours (KNN), support vector machine (SVM), logistic regression (LR), and decision tree (DT) classifiers to predict chronic kidney disease. Finally, they have concluded that SVM has the highest accuracy of the other classifiers[1].

- A. Salekin and J. Stankovic proposed three classifiers: K-NN, RF and ANN to predict CKD. The researchers have used a dataset with 400 patients from UCI with 24 attributes. They implemented wrapper feature selection to find the attributes that detect this disease with high accuracy[2]. Applying the wrapper feature selection five attributes are selected and used in model construction. Finally, they concluded that they could predict CKD with 98% F1 and 0.11 RMSE using RF. F

- F. Aqlan, R. Markle, and A. Shamsan Proposed predictive models such as Decision Trees, Logistic Regression, Naive Bayes, and Artificial Neural Networks for CKD prediction purposes. The researchers conducted data pre-processing to increase the accuracy of the model and reduce overfitting. The researchers used a dataset with 400 patients from UCI with 24 attributes. Researchers have used and compared six predictive models to predict the presence of CKD. The researchers come up with Random Trees as the best method for the CKD prediction in their study

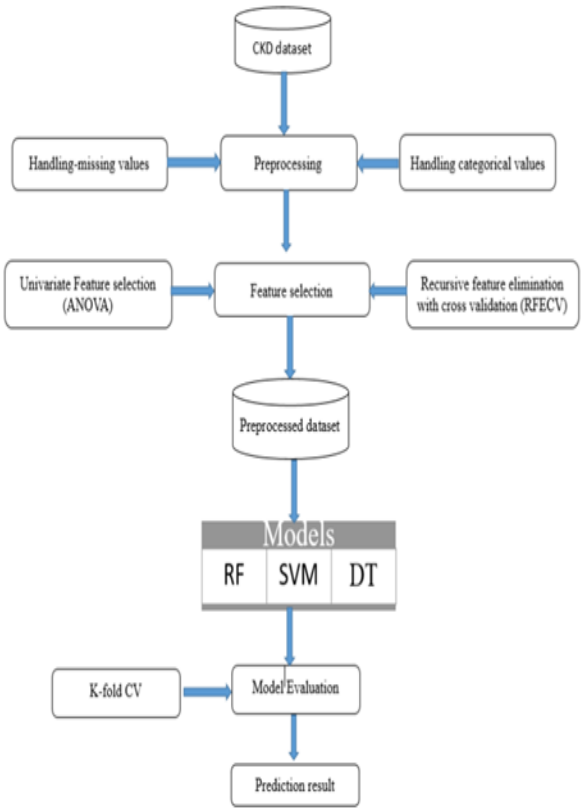
- A. Subas, E. Alickovic, and J. Kevric conducted their study on the Diagnosis of Chronic Kidney

Disease using Random Forest Algorithms. The researchers have used a dataset with 400 patients from UCI with 24 attributes.

- S. Tekale et al. worked on "Prediction of Chronic Kidney Disease Using Machine Learning Algorithm" with a dataset consisting of 400 instances and 14 features. They have used the Decision tree and support vector machine models to implement the prediction system. The researchers pre-processedthe dataset and reduced the number of features from 23 25 to 14. Finally, they come up with SVM as the better model than DT with an accuracy of 96.75%.

- S. Zeynu and S. Patil proposed K-NN, J48, ANN, NB, and SVM classifiers to predict Chronic Kidney Disease. In the research Information gain attributes, an evaluator with ranker search engine and wrapper subset evaluator with the best first engine was used as the feature selection method. The result of their research showed ANN has highest accuracy by using Wrapper feature selection method with Best first search engine feature selection method. J48 and SVM classifiers have an accuracy of 98.75% and 98.25% with Info Gain Attribute Evaluator with a ranker search engine. Building the second method ensemble model by combining the five different classifiers based on a voting algorithm. They showed that the proposed ensemble model achieved 99% accuracy.

- S. Mohammed and T. Beshah conducted their research to develop a self-learning knowledge-based system for diagnosis and treatment of the first three stages of chronic kidney using machine learning. The researchers have used a small number of data and they have developed the prototype, which enables the patient to query KBS to see the delivery of advice. They used decision tree modeling approach in order to model the rules. The researchers conclude that the overall performance of the prototype system registered a 91% accurate result.



Hence, the above reviews indicate that several studies have been conducted on chronic kidney disease prediction using machine-learning techniques. However, the previously conducted research was not concerned with chronic kidney disease severity prediction and the dataset used for previously conducted research is small size, noisy, and not recently collected data. Besides, chronic kidney disease predictions using machine learning were not conducted in Ethiopia. Therefore, in this study the machine-learning model that can predict the severity level of chronic kidney disease was built using a dataset with larger instances than real-world dataset and an accuracy of 99% was achieved using Random-Forest algorithm.

V.APPROACH

CKD prediction using Random Forest algorithm and various comparisons:

A. Decision Tree

Decision Tree (DT) is one of the most popular supervised machine-learning algorithms that can be used for both classification and regression problems. Decision Tree solves the problem of machine learning by transforming the data into a tree representation by

sorting them by feature values. Each node in a decision tree denotes features in an instance to be classified, and each leaf node represents a class label the instances belong to. This model uses a tree structure to split the dataset based on the condition as a predictive model that maps observations about an item to make a decision on the target value of instances.

B. Support Vector Machine

This algorithm is the highest prominent and convenient supervised machine-learning algorithm that can be used for data classification, learning, and prediction. Support Vector Machine builds a set of hyperplanes to classify all input in high dimensional data. In SVM data, classified into two data points in which the hyperplane lies between two classes. SVM creates a discrete hyperplane in the signifier space of the training data and compounds are classified based on the side of the hyperplane located. 9 Hyperplanes are decision boundaries that help separate the data points and support vectors are data points that are closer to the hyperplane and determine the position and orientation of the hyperplane. SVMs were mainly proposed to deal with binary classification, but nowadays many researchers have tried to apply it to multiclass classification because there is a huge amount of data to be classified into more than two classes in today's world. Support Vector Machine (SVM) solves multiclass problems through the two most popular approaches one-versus-rest (one-vs-all) and one-vs-one. In this report, one-versus-rest is used.

C. Random Forest

Random forest is ensemble learning, which is the supervised machine learning algorithm that consists of several collections of decision trees and is used for classification, regression. This model consists of several decision trees and outputs the class target that is the highest voting results of the target output by each tree. Random Forest uses both bagging and random feature selection to build the tree and creates an uncorrelated forest of trees whose prediction by the group is more accurate than that of any individual tree. After it builds the forest, test instances are permeated down through each tree and trees make their respective prediction of class. This is considered as the most efficient algorithm to be used for prediction.

Evaluation Metrics

Prediction model evaluation metrics are required to measure model performance because classifier algorithms do not give the same result. Different performance evaluation metrics such as precision, accuracy, recall, sensitivity, specificity is used in this study with cross-validation.

The followings are the fundamental terms in performance measure:

- True positive (TP): are the condition when both actual value and predicted value are true.
- True negative (TN): are the condition when both the actual value of the data point and the predicted are False.
- False positive (FP): These are the cases when the actual value of the data point was False and the predicted is true.
- False negative (FN): are the cases when the actual value of the data point was true and the predicted is False.

A. Classification Precision

Precision measures the true values correctly predicted from the total predicted values in the actual class. Precision quantifies the ability of the classifiers to not label a negative example as positive. The equation to calculate precision is:

Precision = TP/(TP+FP)

B. Recall

Recall measures the rate of positive values that are correctly classified. Recall answers what proportion of actual Positives is correctly classified. Recall calculated using the equation:

Recall = TP/(TP+FN)

C. F-measure

F-measure is also called F1-score is the harmonic mean between recall and precision. The score is calculated by the equation:

F1 score=2* (Precision*Recall)/(Precision+Recall)

D. Sensitivity

Sensitivity is also called True Positive Rate. Sensitivity is the mean proportion of actual true positives that are correctly identified.

Sensitivity = TP/(TP+FN)

E. Specificity

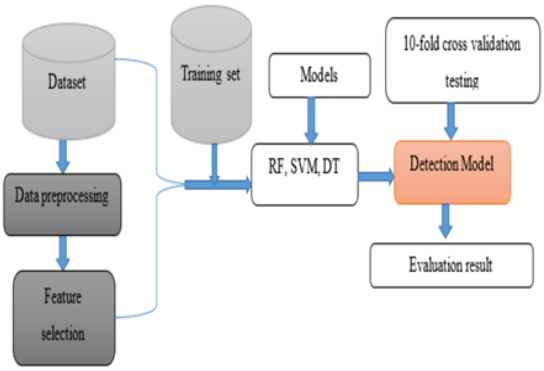
Specificity is also called True Negative Rate. It is used to measure the fraction of negative values that are correctly classified.
Specificity = TN/(TN+FP)

Table 6.3 Performance Result of Binary Classification Models

Models	Performance evaluation metrics				
	Precision	Recall	F1_score	Sensitivity	Specificity
RF	99.9	99.5	99.7	99.5	99.9
SVM	99.1	94.7	96.8	94.7	99.1
DT	99.4	97.6	98.5	97.6	99.4

VI.PROJECT PLANNING

1. Project modelling
This study aims to develop a model that enables us to predict whether the patient has chronic kidney disease or not and its risk level. This study builds a classification model mainly by using machine learning models, RF, SVM, and DT. For creating a predictive model, many records were collected for training and testing using learning algorithms. In this study, the classification process outcome is predicted from a given input or futures relation to class.



2. Models Evaluation and Testing
Model evaluation and testing are critical in our study to test and estimate the performance of machine learning models trained on the chronic kidney disease dataset of both binary class and five class. Since three algorithms were used in this study, model evaluation is very important to compare and select the best model for further prediction. In this study, the effectiveness of the predictive model is evaluated by using 10- fold cross-validation with different performance evaluation metrics appropriate for models; these metrics are accuracy, precision, recall, sensitivity, and specificity as discussed in chapter four

3. Deployment
This study aims to predict chronic kidney disease using three models and recommend the best model to be used for deployment on the server after it is evaluated with the best accuracy and f1 -score. This is done by preparing the HTML form to insert the new input that is connected to the server and selected model that will allow the professionals to diagnose and give the best treatment and advice for the patient fast and correctly.

VII.CONCLUSION

Early prediction is very crucial for both the experts and the patients to prevent and slow down the progress of chronic kidney disease to kidney failure. In this study, the chronic kidney disease prediction using the machine-learning technique is proposed with the deployment of the model to help the experts to diagnose the disease fast. The proposed study also employs the feature selection method to select the most relevant and predictive features. Three machine-learning models RF, SV, DT were used to build the proposed model. The evaluation of the model was done using 10-fold cross-validation. First, the three machine learning algorithms applied to original datasets with all 18 features.

VIII.RESULTS

Applying the models on the original dataset, we have got the highest accuracy with RF scored accuracy of 99.7% for the binary class dataset and lower accuracy of the rest models. SVM and DT produced low performance compared to RF having a scored accuracy of 94.16% and 98.33% respectively, but SVM scored the least accuracy than the rest models. The RF also produced the highest accuracy values than the rest models. Hence, Chronic Kidney disease Predictor is helpful in the healthcare sector and can be deployed soon with the evolution of machine learning.

REFERENCES

[1] C. George, A. Mogueo, I. Okpechi, J. B. Echouffo-Tcheugui, and A. P. Kengne, “chronic

kidney disease in low-income to middle-income countries: The case of increased screening," BMJ Glob. Heal., vol. 2, no. 2, pp. 1–10, 2017.

[2] J. Radhakrishnan and S. Mohan, "KI Reports and World Kidney Day," Kidney Int. Reports, vol. 2, no. 2, pp. 125–126, 2017.

[3] J. W. Stanifer et al., "The epidemiology of chronic kidney disease in sub-Saharan Africa: A systematic review and meta-analysis," Lancet Glob. Heal., vol. 2, no. 3, pp. e174–e181, 2014.

[4] S. Abd Elhafeez, D. Bolignano, G. D'Arrigo, E. Dounousi, G. Tripepi, and C. Zoccali, "Prevalence and burden of chronic kidney disease among the general population and high-risk groups in Africa: A systematic review," BMJ Open, vol. 8, no. 1, 2018.

[5] M. D. Molla et al., "Assessment of serum electrolytes and kidney function test for screening of chronic kidney disease among Ethiopian Public Health Institute staff members, Addis Ababa, Ethiopia," BMC Nephrol., vol. 21, no. 1, p. 494, 2020.

[6] A. Agrawal, H. Agrawal, S. Mittal, and M. Sharma, "Disease Prediction Using Machine Learning," SSRN Electron. J., no. May, pp. 6937–6938, 2018.

[7] S. Mohammed and T. Beshah, "Amharic based knowledge-based system for diagnosis and treatment of chronic kidney disease using machine learning," Int. J. Adv. Comput. Sci. Appl., vol. 9, no. 11, pp. 252–260, 2018.

[8] Y. Amirgaliyev, "Analysis of Chronic Kidney Disease Dataset by Applying Machine Learning Methods," 2018 IEEE 12th Int. Conf. Appl. Inf. Commun. Technol., pp. 1– 4, 2010.

[9] G. Kaur and A. Sharma, "Predict chronic kidney disease using data mining algorithms in hadoop, "International Conference on Inventive Computing and Informatics (ICICI)", Coimbatore,2017.

[10] Y. Saeys, P. Larra, P. Systems, and B.- Ghent, "A review of feature selection techniques in bioinformatics," no. November 2007, 2015.