International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 06 | June - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

PREDICTION OF CHRONIC KIDNEY DISEASE USING MACHINE LEARNING

[1]Miss.Pallavi Patil, [2]Raksha, [3]Saisrijha N Kota, [4] Shreya Patil
[1, 2,3,4] *Department of Computer Science and Engineering,*
Poojya Doddappa Appa College of Engineering, Kalaburagi, Karnataka, India
[1]pallavipatil@pdaengg.com

***Abstract*- Chronic Kidney Disease (CKD) or chronic renal disease is a significant problem that is steadily increasing. The survival time without kidneys is approximately 18 days, leading to a high demand for kidney transplants and dialysis. It is crucial to have reliable techniques for early CKD detection. Machine learning techniques have proven effective in predicting CKD. This study introduces a process for forecasting CKD status using clinical data, which includes data preprocessing, a missing value handling approach employing collaborative filtering, and attribute selection. Among the various machine learning methods examined, the adaboost classifier demonstrated the highest accuracy and minimal bias when predicting CKD status.**

***Keywords*- Chronic Kidney Disease, Machine Learning, Ada boost Classifier, Classification Model.**

I. INTRODUCTION

The kidneys, shaped like beans, are positioned just below the rib cage, one on each side of the spine. On a daily basis, they filter approximately 120 to 150 quarts of blood, resulting in the production of around 1 to 2 quarts of urine. The primary role of the kidneys is to eliminate waste products and excess fluid from the body through urine. They effectively process all the blood in our body about 40 times per day. When chronic renal disease reaches an advanced stage, the body begins to retain harmful amounts of water, electrolytes, and waste substances. CKD indicates that waste has accumulated in the body. This condition is referred to as chronic because the damage occurs gradually over an extended period. Over 14% of the world population suffer from CKD and over 2 million people worldwide currently receive treatment with dialysis or a kidney transplant to stay alive. Diabetes, excessive blood pressure, and coronary heart ailment are most effective three of the various situations which can result in CKD. If one or each of your kidneys aren't operating right, you could have variety of symptoms, inclusive of returned pain, back pain, diarrhea, fever, nosebleeds, rash, and vomiting. The 2 maximum not un usual place ailments that could reason long-time period harm to the kidneys are diabetes and excessive blood pressure. The stages of CKD are mainly based on the measured or estimated glomerular filtration rate (eGFR) which is based on creatinine level, gender, and age. There are five stages of kidney functionality

Stage	Description	GFR(mL/min)
-	At increased risk for CKD	>=90 with risk factors
1	Kidney damage with normal or increased GFR	>=90
2	Mild decrease in GFR	60-89
3	Moderate decrease in GFR	30-59
4	Severe decrease in GFR	15-29
5	Kidney Failure	<15 or dialysis

Table 1: Stages of Chronic Kidney Disease [14]

Fig 1 shows that after the stage 2 of CKD, patient will suffer symptoms and will get to know about the reducing of kidney functionality. During initial stages, CKD has no symptoms. Therefore, early detection of CKD can reduce a patient's chances of CKD. With advances in machine learning and artificial intelligence, several classifiers and clustering algorithms are used in order to achieve this goal.

II. LITERATURE SURVEY

Summary of research papers survey are listed down, the observations made in each paper are summarized below the respective paper and these observations are used to improve the overall system.

1. Prathiba Devishri S, Ragin O R, Anisha GS, "Comparitive Study of Classification Algorithms in CKD", International Journal of Recent Technology and Engineering 2019.

This paper proposes feature selection approach is suitable for CKD predictions and examined decision stump, Rep tree, IBK, K-star, SGD and SMO classifier using WEKA. Accuracy measures used to compare classifiers are Recall. F-measure and prediction by implementing on WEKA, they got 87.3 accuracy. Decision stump and Rep tree algorithms gave better results with less error rate.

2. Sujata Drall, Gurdeep Singh Drall, Sugandha Singh, Bharat Bhushan Naib "Chronic Kidney Disease Prediction using Machine Learning", International Journal of Management, Technology and Engineering,2018.

This paper proposes an innovative method by conducting Data Mining, Machine Learning and different classification algorithms which is focused on predicting CKD status of a sick person with high accuracy. they have used 5 CKD attributes out of 25 and 2 classification algorithms KNN and Naive Bayes Classifier for portending CKD of a patient. KNN Classifier predicted CKD with an accuracy of 100% ,whereas Naive Bayes Classifier predicted with an accuracy of 96.25%.

3. El –Houssainy A. Rady, Ayman S.Anwar "Prediction of Kidney Disease stages using Data Mining Algorithms". Informatics in Medicine Unlocked (IMU 100178), 2019.

This paper proposes Effective Data Mining Strategies is appeared to uncover besides concentrate concealed data the outcomes of PNN, MLP, SVM and RBF procedures have been associated PNN process gives the maximum accuracy of 99.7%as related with all other algorithm consequences.

4. Siddheshwar Tekale, Pranjal Shingavi, Sukanya Wandhekar, "prediction of Chronic Kidney Disease using Machine Learning Algorithms" International Journal of Advance Research in computer and communication Engineering,2018

This paper proposes 14 dissimilar characteristics correlated to CKD sick person and anticipated accuracy of several ML procedures alike Decision tree and Support Vector Machine Only 14 of the 25 features were employed in model. Only 14 of the 25 features were employed in the predictive model. It is observed from the results of study that decision tree algorithms gives the accuracy of 91.75% and SVM renders accuracy of 96.75%.

5. S. Gopika, Dr.M.Vanitha, "Machine learning Approach of chronic kidney disease prediction using clustering Techniques" ,International Journal of Innovative Research in Science, Engineering and Technology ,2017.

This paper proposes a method of CKD prediction with clustering method. The main objective is to determine the kidney function failure by using clustering algorithm. The experimental outcome revealed that the Fuzzy C means algorithm renders superior results and its accuracy 89%.

6. Faisal Aqlan, Ryan Markle, Abdulrahman Shamsan, "Data Mining for chronic kidney disease prediction",67thAnnual Conference and expo of the institute of Industrial Engineers, United States 2017

This paper proposes a DST that assist in the analysis of CKD. Data Mining and analytics methods can be designed for detecting CKD by conducting chronological patient's report and analysis proceedings. DT, LR, NB, and ANN were intended for detecting CKD. Random Trees gives 100% accuracy.

7. Sahil Sharma, Vinod Sharma and Atul Sharma, "Performance Based Evaluation of Various Machine Learning Classification Techniques for CKD Diagnosis". International Journal of Modern Computer Science (IJMCS),2016.

This paper proposes various machine learning algorithms for CKD. 400 cases and 24 features are used for the research. The Outcomes demonstrate that DT did the best results with the accuracy of 98.6%, sensitivity of 0.9720, precision of 1 and specificity of 1.

8. Asif Salekin, john stankovic "Detection of chronic Kidne Disease and Selecting Important predictive Attributes" IEEE EXPLORE 2016.

This paper proposes Evaluation of three classifiers to detect CKD K-nearest neighbor random forest Neural Networks they achieved a prediction accuracy of 0.993. they additionally performed feature assortment to choose the utmost appropriate features for recognizing CKD and flourishing them consistent with their certainty.

9. J. Snegha, “Chronic Kidney Disease Prediction using Data Mining”, International Conference on Emerging Trends, 2020.

This paper proposes a system that uses various data mining techniques like Random Forest algorithm and Back propagation neural Network. Here they compare both of the algorithm and found that Back Propagation algorithm gives the best result as it uses the supervised learning network called feedforward neural network.

10. Devika R, Sai Vaishnavi A, Subramaniyaswamy V (2019) Comparative Study of Classifier for Chronic Kidney Disease Prediction Using Naive Bayes, KNN and Random Forest. “2019 3rd International Conference on Computing Methodologies and Communication (ICCMC).”

In this paper author published Comparative Study of Classifier for Chronic Kidney Disease Prediction Using Naïve Bayes, KNN and Random Forest Chronic Kidney disease defines constrains which affects your kidneys and reduces your potential to stay healthy. Machine learning is an important task as it benefits many applications, varied knowledge mining classification approaches and machine learning algorithms are applied for prediction of chronic diseases. Therefore, this paper examines the performance of Naive Bayes, K-Nearest Neighbor (KNN) and Random Forest classifier on the basis of its accuracy, preciseness and execution time for CKD prediction Finally, the outcome after conducted research is that the performance of Random Forest classifier is finest than Naive Bayes and KN

11. Baisakhi Chakraborty, “Development of Chronic Kidney Disease Prediction Using Machine Learning”, International Conference on Intelligent Data Communication Technologies, 2019

In this paper author proposed development of CKD prediction system using machine learning techniques such as K-Nearest Neighbor, Logistic Regression, Decision Tree, Random Forest, Naïve Bayes, Support Vector Machine and Multi-Layer Perceptron Algorithm. These are applied and their performance are compared to the accuracy, precision, and recall results. Finally, Random forest is chosen to implement this system.

12. Arif-Ul-Islam and S. H. Ripon, "Rule Induction and Prediction of Chronic Kidney Disease Using Boosting Classifiers, Ant-Miner and J48 Decision Tree," 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), Cox'sBazar, Bangladesh, 2019, pp. 1-6

In this paper author proposed a system in which prediction of disease is done using Boosting Classifiers, Ant-Miner and J48 Decision Tree. The aim of this paper is two fold that is, analyzing the performance of boosting algorithms for detecting CKD and deriving rules illustrating relationships among the attributes of CKD. Experimental results prove that the performance of AdaBoost was less that of LogitBoost by a fraction.

13. Amirgaliyev Y, Shamilulu S, Serek A (2019) Analysis of Chronic Kidney Disease Dataset by Applying Machine Learning Methods. 2019 IEEE 12th International Conference on Application of Information and Communication Technologies (AICT).

In this paper author published Analysis of Chronic Kidney Disease Dataset by Applying Machine Learning Methods prof. Amirgaliyev entitled Analysis of Chronic Kidney Disease Dataset by Applying Machine Learning Methods. Currently, there are many people in the world suffering from chronic kidney diseases worldwide. Due to the several risk factors like food, environment and living standards many people get diseases suddenly without understanding of their condition. In this research study, the effects of using clinical features to classify patients with chronic kidney disease by using support vector machines algorithm is investigated. The chronic kidney disease data set is based on clinical history, physical examinations, and laboratory tests.

14. S.Ramya, Dr. N.Radha, "Diagnosis of Chronic Kidney Disease Using Machine Learning Algorithms," Proc. International Journal of Innovative Research in Computer and Communication Engineering,Vol. 4, Issue 1, January 2016

In this paper author worked on diagnosis time and improvement of diagnosis accuracy using different classification algorithms of machine learning. The proposed work deals with classification of different stages of CKD according to its gravity. By analysing different algorithms like Basic Propagation Neural Network, RBF and RF. The analysis results indicates that RBF algorithm gives better results than the other classifiers and produces 85.3% accuracy

15. Guneet Kaur, "Predict Chronic Kidney Disease using Data Mining in Hadoop, International Conference on Inventive Computing and Informatics, 2017.

This paper proposes a system for predicting the CKD using Data Mining Algorithms in Hadoop. They use two data mining classifiers like KNN and SVM. Here the predictive analysis is performed based upon the manually selected data columns. SVM classifier gives the best accuracy than KNN in this system.

III. METHODOLOGY

The proposed methodology consists of 3 key steps: Data preprocessing, models training and model selection

A. Dataset: We have utilized the publicly accessible CKD Dataset from the UCI repository for our analysis. This dataset consists of 400 samples belonging to two distinct classes. Among the 25 attributes present, 11 are numeric, 13 are nominal, and one attribute represents the class. The dataset contains some missing values. The dataset provides information about the patients, including their age, blood pressure, specific gravity, albumin, sugar, red blood cells, and other relevant factors

No.	Attributes	Description	Type
1	Al	Albumin	Nominal
2	Ane	Anemia	Nominal
3	appet	Appetite	Nominal
4	Ba	Bacteria	Nominal
5	Cad	Coronary artery disease	Nominal
6	Dm	diabetes mellitus	Nominal
7	Htn	Hypertension	Nominal
8	Pc	pus cell	Nominal
9	Pcc	pus cell clumps	Nominal
10	Pe	pedal edema	Nominal
11	Rbc	red blood cells	Nominal
12	Sg	specific gravity	Nominal
13	Su	Sugar	Nominal
14	Age	Age	Numeric
15	Bgr	blood glucoses	Numeric
16	Bp	blood pressure	Numeric
17	Bu	blood urea	Numeric
18	Hemo	Haemoglobin	Numeric
19	Pcv	packed cell volume	Numeric
20	Pot	Potassium	Numeric
21	Rc	red blood cell count	Numeric
22	Sc	serum creatinine	Numeric
23	Sod	Sodium	Numeric
24	Wbcc	white blood cell	Numeric
25	Class	Class	Nominal

Table.2 List of attributes present in the CKD dataset

CKD is caused due to diabetes and high blood pressure. Due to Diabetes our many organs get affected and it will be followed by high blood sugar. So it is important to predict the disease as early as possible

Fig 1: Proposed workflow

B. Data pre-processing: Data Pe-Processing is that stage where the data that is distorted, or encoded is brought to such a state that the machine can easily analyze it. A dataset can be observed as a group of data objects. Data objects are labelled by a number of features, that ensures the basic features of an object, such as the mass of a physical object or the time at which an event ensured. In the dataset there may be missing values, they can either eliminated or estimated.

C. Feature Selection: Feature Selection is the method where we computationally select the features which contribute most to our prediction variable or output.

D. Model Training

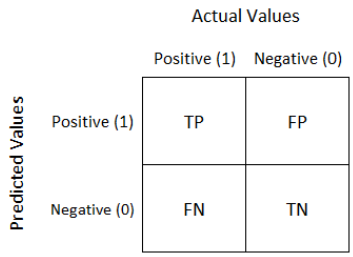
Model training is a process in which a machine learning (ML) algorithm is fed with sufficient training data to learn from.

E. Model selection

It is the process of choosing one among many candidate models for a predicting problems

IV. Result and Discussion

The metrics provided below gives us information on the quality of the outcomes that we get in this study. A confusion matrix helps us with this by describing the performance of the classifier



Precision: Precision or positive predictive value here is the ratio of all patients actually with CKD to all the patients predicted with CKD (true positive and false positive).

$$\text{Precision} = \frac{TP}{FP + TP}$$

Recall: It is also known as sensitivity and it is the ratio of actual number of CKD patients that are correctly identified to the total no of patients with CKD.

$$\text{Recall} = \frac{TP}{FN + TP}$$

F- Measure: It measures the accuracy of the test. It is the harmonic mean between precision and recall.

$$\text{F-Measure} = \frac{2 * \text{Recall} * \text{Precision}}{\text{Recall} + \text{precision}}$$

Accuracy: It is the ratio of correctly predicted output cases to all the cases present in the data set.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN}$$

Support: Support is the correct number of outcomes or responses that are present in each class of the predicted outcome.

Models	Accuracy	Precision	Recall	f1-score
Decision Tree	0.95	0.94	0.95	0.95
Ada boost	0.98	1.0	0.96	0.98
KNN	0.64	0.55	0.62	0.58
Cat boost	0.97	0.98	0.96	0.97
Xg boost	0.96	0.96	0.96	0.96

Table 3: Performance comparison of different models (not ckd)

Models	Accuracy	Precision	Recall	f1-score
Decision Tree	0.95	0.97	0.96	0.97
Ada boost	0.98	0.97	1.00	0.99
KNN	0.64	0.72	0.65	0.69
Cat boost	0.97	0.97	0.99	0.98
Xg boost	0.96	0.97	0.97	0.97

Table 4: Performance comparison of different models (ckd)

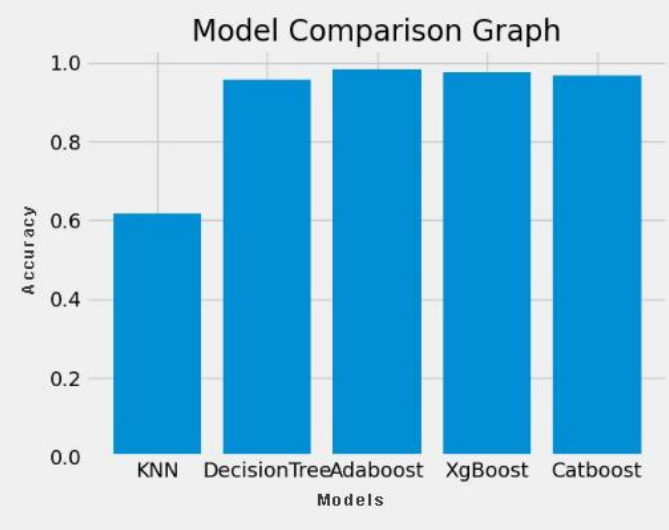


Fig 2: Model comparison graph

VI. Conclusion

This project focuses on forecasting Chronic Kidney Disease (CKD) in individuals. We examined 14 distinct factors associated with CKD patients and evaluated the accuracy of various machine learning techniques such as Decision tree, ada boost, K-NN, catboost, Xgboost. The system offers the benefit of quicker prediction, enabling doctors to initiate early treatment for CKD patients. Additionally, it aids in diagnosing a larger number of patients in a shorter timeframe.

References

[1].PrathibaDevishri S, Ragin O R, Anisha GS, "Comparitive Study of Classification Algorithms in CKD", International Journal of Recent Technology and Engineering 2019

[2]. SujataDrall, Gurdeep Singh Drall, Sugandha Singh, Bharat Bhushan Naib "Chronic Kidney Disease Prediction using Machine Learning" ,International Journal of Management Technology and Engineering.2018

[3]. El –HoussainyA.Rady ,AymanS.Anwar "Prediction of Kidney Disease stages using Data Mining Algorithms". Informatics in Medicine Unlocked(IMU 100178), 2019

[4] SiddheshwarTekale, PranjalShingavi, sukanyaWandhekar ,"prediction of Chronic Kidney Disease using Machine Learning Algorithms" International Journal of Advance Research in computer and communication Engineering,2018

[5] S. Gopika, Dr.M.Vanitha, "Machine learning Approach of chronic kidney disease prediction using clustering Techniques", International Journal of Innovative Research in Science, Engineering and Technology ,2017

FaisalAqlan, Ryan Markle, Abdulrahman Shamsan, "Data Mining for chronic kidney disease prediction",67thAnnual Conference and expo of the institute of Industrial Engineers,United States 2017

[7] Sahil Sharma ,Vinod Sharma and Atul Sharma, "Performance Based Evaluation of Various Machine Learning Classification Techniques for CKD Diagnosis". International Journal of Modern Computer Science (IJMCS), 2016.

[8]Asif Salekin, john stankovic "Detection of chronic Kidney Disease and Selecting Important predictive Atributes" IEEE EXPLORE 2016

[9] J. Snegha, "Chronic Kidney Disease Prediction using Data Mining", International Conference on Emerging Trends, 2020

[10] Devika R, Sai Vaishnavi A, Subramaniyaswamy V (2019) Comparative Study of Classifier for Chronic Kidney Disease Prediction Using Naive Bayes, KNN and Random Forest. "2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)."

[11] Baisakhi Chakraborty, "Development of Chronic Kidney Disease Prediction Using Machine Learning", International Conference on Intelligent Data Communication Technologies, 2019

[12] Arif-Ul-Islam and S. H. Ripon, "Rule Induction and Prediction of Chronic Kidney Disease Using Boosting Classifiers, Ant-Miner and J48 Decision Tree," 2019 International Conference on Electrical, Computer and Communication Engineering

(ECCE), Cox'sBazar, Bangladesh, 2019, pp. 1-6

[13] Amirgaliyev Y, Shamilulu S, Serek A (2019) Analysis of Chronic Kidney Disease Dataset by Applying Machine Learning Methods. 2019 IEEE 12th International Conference on Application of Information and Communication Technologies (AICT)

[14] S.Ramya, Dr. N.Radha, "Diagnosis of Chronic Kidney Disease Using Machine Learning Algorithms," Proc.

International Journal of Innovative Research in Computer and Communication Engineering,Vol. 4, Issue 1, January 2016.

[15] Guneet Kaur, "Predict Chronic Kidney Disease using Data Mining in Hadoop, International Conference on Inventive Computing and Informatics, 2017.

[16] A. J. Aljaaf, D. Al-Jumeily, H. M. Haglan, M. Alloghani, T. Baker, A. J. Hussain, and J. Mustafina, "Early prediction of chronic kidney disease using machine learning supported by predictive analytics," in 2018 IEEE Congress on Evolutionary Computation (CEC). IEEE, 2018, pp. 1–9..

[17] P. Yildirim, "Chronic kidney disease prediction on imbalanced data by multilayer perceptron: Chronic kidney disease prediction," 2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC), vol. 02, pp. 193–198, 2017..

[18] J. Xiao, R. Ding, X. Xu, H. Guan, X. Feng, T. Sun, S. Zhu, and Z. Ye, "Comparison and development of machine learning tools in the prediction of chronic kidney disease progression," Journal of translational medicine, vol. 17, no. 1, p. 119, 2019.

[19] S. Nair, S. V. O'Brien, K. Hayden, B. Pandya, P. J. G. Lisboa, K. J. Hardy, and J. P. H. Wilding, "Effect of a Cooked Meat Meal on Serum Creatinine and Estimated Glomerular Filtration Rate in Diabetes-Related Kidney Disease," Diabetes Care, vol. 37, no. 2, pp. 483–487, Feb. 2014

[20] Vijayarani, S. Dhayanand et al., "Data mining classification algorithms for kidney disease prediction," International Journal on Cybernetics & Informatics (IJCI), vol. 4, no. 4, pp. 13–25, 2015