IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.134160

PREDICTION OF CHRONIC KIDNEY DISEASE USING MACHINE LEARNING

Supriya G1, Swathi J2, Chandrika K3, Vamsi Krishna V4

Student, Computer Science and Engineering & Business Systems, RGMCET, Nandyala, India1

Student, Computer Science and Engineering & Business Systems, RGMCET, Nandyala, India2

Student, Computer Science and Engineering & Business Systems, RGMCET, Nandyala, India3

Student, Computer Science and Engineering & Business Systems, RGMCET, Nandyala, India4

Abstract -Early diagnosis and understanding of Chronic Kidney Disease (CKD) are critical for effective treatment planning, as CKD profoundly impacts kidney function, leading to complications like bone and mineral concerns, low blood pressure, acid-base imbalances, poor nutrition, and neurological disorders. This study explores the application of machine learning (ML) algorithms and various data mining classification methods to predict and diagnose CKD. Leveraging a dataset with 21 features from the UCI Repository, algorithms including Logistic Regression, Decision Tree, SVM, Bagging, Adaboost, Voting Classifier, KNN, Xgboost Gradient Boosting, and Random Forest were employed. Notably, Random Forest exhibited remarkable accuracy at 98.75%. The findings underscore the potential of ML in enhancing CKD identification. This research contributes to the growing body of knowledge on utilizing advanced analytics in healthcare. The 98.75% accuracy achieved by Random Forest emphasizes its efficacy in early CKD detection, offering valuable insights for improved patient care .

Keywords:- Chronic kidney disease, Machinelearning, XgBoost classifier, Classification model.

I. INTRODUCTION

A major worldwide health concern, CKD is defined Chronic Kidney Disease (CKD) represents a significant global health challenge, impacting the vital functionsof the kidneys and giving rise to a myriad of omplications, ranging from bone and mineralconcerns to neurological disorders. Early detection and understanding of CKD are paramount for devising effective treatment strategies that can prevent the exacerbation of patients' health. In this context, the integration of machine learning (ML) algorithms and data mining classification methods emerges as a promising approach for predictive analysis and accurate diagnosis of CKD.

This research endeavors to contribute to the field by leveraging a comprehensive dataset from the UCI Repository, comprising 400 instances with 21 distinct features. Through the application of diverse ML algorithms, including Logistic Regression, Decision Tree, SVM, Bagging, Adaboost, Voting Classifier, KNN, Xgboost Gradient Boosting, and Random Forest, the study aims to unravel patterns and correlations KNN, Xgboost Gradient Boosting, and Random Forest,

the study aims to unravel patterns and correlations within the data to enhance CKD identification. The focal point of this investigation lies in the exceptional accuracy demonstrated by the Random Forest algorithm, reaching an impressive 98.75%, thereby showcasing its potential as a robust tool in the early detection and management of CKD.

At the forefront of medical imaging, image segmentation emerges as a critical step in image processing, wielding significant influence over the success of subsequent stages [2]. In the realm ofmedical image processing, particularly in the detection of tumors or lesions, efficient machine vision and precise segmentation are paramount for accurate diagnosis and treatment planning.

Evaluate the performance of various machine learning algorithms, including Logistic Regression, Decision Tree, SVM, Bagging, Adaboost, Voting Classifier, KNN, Xgboost Gradient Boosting, and Random Forest, for predicting and diagnosing Chronic Kidney Disease (CKD).Utilize a dataset with 21 features from the UCI Repository to assess the accuracy of the mentioned algorithms in early CKD detection.Contribute valuable insights to the field of healthcare analytics by emphasizing the potential of machine learning, particularly highlighting the remarkable 98.75% accuracy achieved by the Random Forest algorithm, and its implications for improving patient care.

DOI: 10.17148/IJARCCE.2024.134160

The purpose of this study is to assess the effectiveness of machine learning algorithms in predicting and diagnosing Chronic Kidney Disease (CKD). By employing various classification methods on a dataset with 21 features, the research aims to highlight the potential of ML, with a focus on the notable 98.75% accuracy achieved by the Random Forest algorithm for early CKD detection. The study contributes toadvancing healthcare analytics and improving patient care.

The lack of efficient and accurate methods for early Chronic Kidney Disease (CKD) diagnosis poses a significant healthcare challenge. This study addresses the need for improved diagnostic tools by evaluating machine learning algorithms on a comprehensive dataset. The focus is on identifying an algorithm with high accuracy, emphasizing the potential of advancedanalytics in enhancing CKD detection for better patient outcomes.

II . LITERATURE SURVEY

The feasibility of two in-house fuzzy classifiers, fuzzy rule-building expert system (FuRES) and fuzzy optimal associative memory (FOAM), for diagnosis of patients with chronic kidney disease (CKD) was investigated. A linear.DA), was used for comparison. The CKD data used in this work were taken from the UCI Machine Learning Repository.
Composite datasets were created by adding different levels of proportional noise to evaluate the robustness of the two fuzzy approaches. Firstly, 11 levels of proportional noises were added to each numeric attribute of the training and prediction sets one after another, and then these simulated training and prediction sets were combined in pairs. Thus, a grid with 121 groups of simulated data was generated, and classification rates for these 121 pairs were compared. Secondly, the performances of two fuzzy classifiers using the simulated datasets, in which 11 levels of noise were randomly distributed to each numeric attribute, were compared and the average prediction rates of FuRES and FOAM were
98.1 ± 0.5% and 97.2 ± 1.2%, respectively, with 200bootstrap Latin partitions. The PLS-DA can give
94.3 ± 0.8% with the identical evaluation. Confluentdatasets comprised of the original and modified datasets were also used to evaluate FuRES, FOAM, and PLS-DA classification models. The average prediction rates of FuRES and FOAM obtained from 200 bootstrapped evaluations were 99.2 ± 0.3% and 99.0 ± 0.3%. PLS-DA yields slightly worse accuracy with 95.9 ± 0.6%. The results demonstrate that both FuRES and FOAM perform well on the identification of CKD patients, while FuRES is more robust than FOAM. These two fuzzyclassifiers are useful tools for the diagnosis of CKD patients with satisfactory robustness, and can also beused for other kinds of patients.

Chronic kidney disease (CKD) is a global public health problem, affecting approximately 10% of the population worldwide. Yet, there is little direct evidence on how CKD can be diagnosed in a systematic and automatic manner. This paper investigates how CKD can be diagnosed by using machine learning (ML) techniques. ML algorithms have been a driving force in detection of abnormalities in different physiological data, and are, with a great success, employed in different classification tasks. In the present study, a number of different ML classifiers are experimentally validated to a real data set, taken from the UCI Machine Learning Repository, and our findings are compared with the findings reported in the recent literature. The results are quantitatively and qualitatively discussed and our findings reveal that the random forest (RF) classifier achieves the near-optimal performances on the identification of CKD subjects. Hence, we show that ML algorithms serve important function in diagnosis of CKD, with satisfactory robustness, and our findings suggest that RF can alsobe utilized for the diagnosis of similar diseases.

The prevalence of chronic kidney disease is high in developing countries. However, no national survey of chronic kidney disease has been done incorporating both estimated glomerular filtration rate (eGFR) and albuminuria in a developing country with the economic diversity of China. We aimed to measure the prevalence of chronic kidney disease in China withsuch a survey.

Methods: We did a cross-sectional survey of a nationally representative sample of Chinese adults. Chronic kidney disease was defined as eGFR less than
60 mL/min per 1·73 m(2) or the presence of albuminuria. Participants completed a lifestyle and medical history questionnaire and had their blood pressure measured, and blood and urine samples taken. Serum creatinine was measured and used to estimate glomerular filtration rate. Urinary albumin and creatinine were tested to assess albuminuria. The crude and adjusted prevalence of indicators of kidney damage were calculated and factors associated with the presence of chronic kidney disease analysed by logistic regression.

Predictive models built using temporal data in electronic health records (EHRs) can potentially play a major role in improving management of chronic diseases. However, these data present a multitude of technical challenges, including

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.134160

irregular sampling of data and varying length of available patient history. In this paper, we describe and evaluate three different approaches that use machine learning to build predictive models using temporal EHR data of a patient. The first approach is a commonly used non- temporal approach that aggregates values of the predictors in the patient's medical history. The other two approaches exploit the temporal dynamics of the data. The two temporal approaches vary in how they model temporal information and handle missing data. Using data from the EHR of Mount Sinai MedicalCenter, we learned and evaluated the models in the context of predicting loss of estimated glomerular filtration rate (eGFR), the most common assessment of kidney function. Our results show that incorporating temporal information in patient's medical history can lead to better prediction of loss of kidney function. They also demonstrate that exactly how this information is incorporated is important. In particular, our results demonstrate that the relative importance of different predictors varies over time, and that using multi-task learning to account for this is an appropriate way to robustly capture the temporal dynamics in EHR data. Using a case study, we also demonstrate how the multi-task learning based model can yield predictive models with better performance for identifying patients at high risk of short-term loss of kidney function.

Background and aims: One strategy to prevent and managechronic kidney disease (CKD) is to offer screening programs. The aim of this study was to determine the percentage prevalence and risk factors of CKD in a screening program performed in an adult general population. This is a cross-sectional study. Six- hundred ten adults (73% women, age 51 ± 14 years) without previously known CKD were evaluated. Participants were subjected to a questionnaire, blood pressure measurement and anthropometry. Glomerular filtration rate estimated by CKD-EPI formula and urine tested with albuminuria dipstick.

Results: More than 50% of subjects reported family antecedents of diabetes mellitus (DM), hypertension and obesity, and 30% of CKD. DM was self-reported in 19% and hypertension in 29%. During screening, overweight/obesity was found in 75%; women had a higher frequency of obesity (41 vs. 34%) and high-risk abdominal waist circumference (87 vs. 75%) than men. Hypertension (both self-reported and diagnosed in screening) was more frequent in men (49%) than in women (38%). CKD was found in 14.7%: G1, 5.9%; G2, 4.5%; G3a, 2.6%; G3b, 1.1%, G4, 0.3%; and G5, 0.3%. Glomerular filtration rate was mildly/moderately reduced in 2.6%, moderately/severely reduced in 1.1%, and severely reduced in <1%. Abnormal albuminuria was found in 13%. CKD was predicted by DM, hypertension and male gender.

Hodneland et al. utilized image registration to detect renal morphologic changes. Vasquez-Morales et al. established a classifier based on neural network using large-scale CKD data, and the accuracy of the model on their test data was 95%. In addition, most of the previous studies utilized the CKD data set that was obtained from the UCI machine learning repository. Chen et al. used k-nearest neighbor (KNN), support vector machine (SVM) and soft independent modeling of class analogy to diagnose CKD, KNN and SVM achieved the highest accuracy of 99.7%. In addition, they used fuzzy rule-building expert system, fuzzy optimal associative memory and partial least squares discriminant analysis to diagnose CKD, and the range of accuracy in those models was 95.5%-99.6%. Their studies have achieved good results in the diagnosis of CKD.

Most of them suffering from either the method used to impute missing values has a limited application range or relatively low accuracy. In the above models, the mean imputation is used to fill in the missing values and it depends on the diagnostic categories of the samples. As a result, their method could not be used

when the diagnostic results of the samples are unknown. In reality, patients might miss some measurements for various reasons before diagnosing. addition, for missing values in categorical variables, data obtained using mean imputation might have a large deviation from the actual values.

III. METHODOLOGY

a) Proposed work:

The proposed system is designed with the primary objective of precise and timely identification of Chronic Kidney Disease (CKD) to prevent the deterioration of patients' health. Leveraging a dataset sourced from the UCI Repository, the system incorporates 400 instances, each characterized by 21 features.

Employing a diverse set of machine learning and data mining techniques, including Logistic Regression, Decision Tree, SVM, Bagging, Adaboost, Voting Classifier (Adaboost + DT + Bagging), KNN, Xgboost Gradient Boosting, and Random Forest, the system aims to comprehensively analyze and classify the dataset. By harnessing the strengths of

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.134160

various algorithms, the proposed system endeavors to enhance accuracy in CKD prediction. This multi- faceted approach allows for a robust evaluation of patient data, contributing to the system's effectiveness in aiding healthcare professionals in theearly detection and management of CKD, therebyimproving overall patient outcomes.

The proposed system is designed with the primary objective of precise and timely identification of Chronic Kidney Disease (CKD) to prevent the deterioration of patients' health. Leveraging a dataset sourced from the UCI Repository, the system incorporates 400 instances, each characterized by 21 features. Employing a diverse set of machine learning and data mining techniques, including Logistic Regression, Decision Tree, SVM, Bagging, Adaboost, Voting Classifier (Adaboost + DT + Bagging), KNN, Xgboost Gradient Boosting, and Random Forest, the system aims to comprehensively analyze and classify the dataset. By harnessing the strengths of various algorithms, the proposed system endeavors to enhance accuracy in CKD prediction. This multi- faceted approach allows for a robust evaluation of patient data, contributing to the system's effectiveness in aiding healthcare professionals in the early detection and management of CKD, therebyimproving overall patient outcomes.

b) System Architecture:

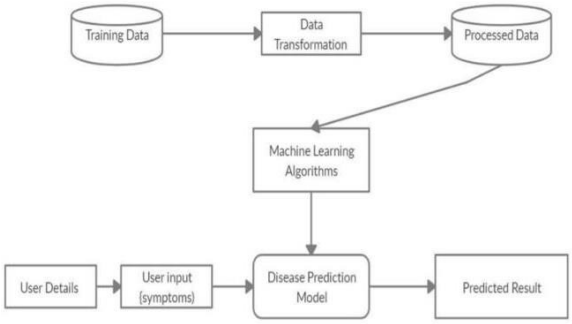
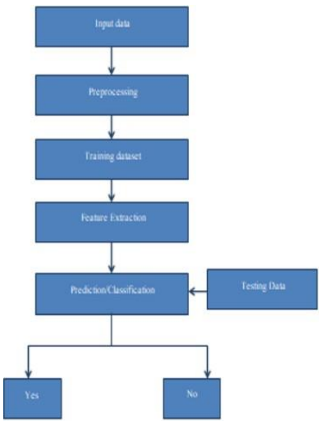


Fig 1:System Architecture

c) Data flow diagram:

1. The DFD is also called as bubble chart. It is a simple graphical formalism that can be used to represent a system in terms of input data to the system, various processing carried out on this data, and the output data is generated by this system.
2. DFD shows how the information moves through the system and how it is modified by a series of transformations. It is a graphical technique that depicts information flow and the transformations that are applied as data moves from input to output.
3. DFD is also known as bubble chart. A DFD may be used to represent a system at any level of abstraction. DFD may be partitioned into levels that represent increasing information flow and functional detail.



IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.134160

UML Diagrams:

UML stands for Unified Modeling Language. UML is astandardized general-purpose modeling language in the field of object-oriented software engineering.
The standard is managed, and was created by, the Object Management Group.The goal is for UML to become a common language for creating models of object oriented computer software. In its current form UML is comprised of two major components: a Meta-model and a notation. In the future, some form of method or process may also be added to; or associated with, UML.The Unified Modeling Language is a standard language for specifying, Visualization, Constructing and documenting the artifacts of software system, as well as for business modeling and other non-software systems.The UML represents a collection of best engineering practices that have proven successful in the modeling of large and complex systems.The UML is a very important part of developing objects oriented software and thesoftware development process. The UML uses mostly graphical notations to express the design of software projects.

The Primary goals in the design of the UML are asfollows:

1. Provide users a ready-to-use, expressive visual modeling Language so that they can develop and exchange meaningful models.
2. Provide extendibility and specializationmechanisms to extend the core concepts.
3. Be independent of particular programming languages and development process.
4. Provide a formal basis for understanding themodeling language.
5. Encourage the growth of OO tools market.
6. Support higher level development concepts such as collaborations, frameworks, patternsand components.
7. Integrate best practices.

A use case diagram in the Unified Modeling Language(UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.

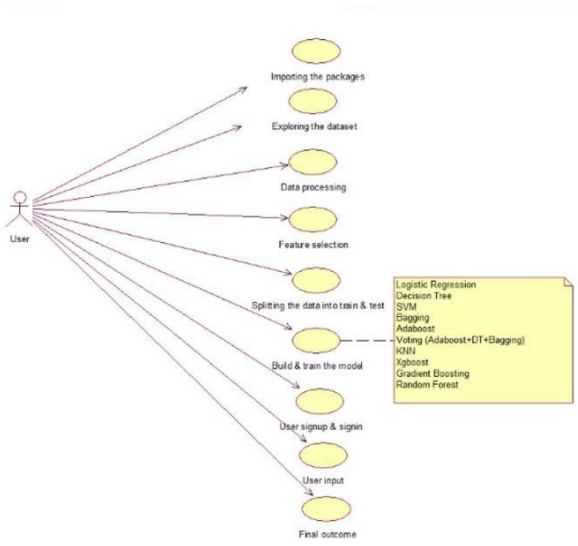


Fig 3. Use case diagram

d) Data Collection:

This is the first real step towards the real development of a machine learning model, collecting data. This is a critical step that will cascade in how good the model will be, the more and better data that we get, the better our model will perform.There are several techniques to collect the data, like web scraping, manual interventions and etc.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024
DOI: 10.17148/IJARCCE.2024.134160

e) Data Preparation:

we will transform the data. By getting rid of missing data and removing some columns. First we will create a list of column names that we want to keep or retain.Next we drop or remove all columns except forthe columns that we want to retain.Finally we drop orremove the rows that have missing values from the data set.Split into training and evaluation sets Model Selection: we will transform the data. By getting rid of missing data and removing some columns. First we will create a list of column names that we want to keep or retain.Next we drop or remove all columns except forthe columns that we want to retain.Finally we drop orremove the rows that have missing values from the data set.Split into training and evaluation sets.

f) Saving the Trained Model:

Once you're confident enough to take your trained and tested model into the production-ready environment, the first step is to save it into a .h5 or .pkl file using a library like pickle.Make sure youhave pickleinstalled in your environment.Next,
let's import the module and dump the model into .pklfile

g) Algorithms:

Logistic Regression: Logistic regression is a statistical analysis method to predict a binary outcome, such as yes or no, based on prior observations of a data set. A logistic regression model predicts a dependent data variable by analyzing the relationship between one ormore existing independent variables.

Decision Tree: A decision tree algorithm is a machine learning algorithm that uses a decision tree to make predictions. It follows a tree-like model of decisions and their possible consequences. The algorithm works by recursively splitting the data into subsets based on the most significant feature at each node ofthe tree.

SVM: A support vector machine (SVM) is a machine learning algorithm that uses supervised learning models to solve complex classification, regression, and outlier detection problems by performing optimal data transformations that determine boundaries between data points based on predefinedclasses, labels, or outputs.
Bagging: Bagged Decision Trees, or Bagging, is an ensemble learning method that combines multiple decision tree classifiers. It reduces overfitting andenhances predictive performance by training on various subsets of the dataset.

Adaboost: AdaBoost is a boosting algorithm that creates a strong classifier by combining weak classifiers. A weak classifier is simply a classifier that performs poorly, but better than random guessing. The algorithm works by weighting the weak classifiersso that they vote with more importance.

Voting Classifier (Adaboost + DT + Bagging): The Voting Classifier combines AdaBoost, Decision Trees (DT), and Bagging algorithms to form an ensemble model. It leverages their diverse strengths, allowing each classifier to vote on predictions, resulting in an improved and more robust classification model.

KNN: The k-nearest neighbors algorithm, also known as KNN or k-NN, is a non-parametric, supervised learning classifier, which uses proximity to make classifications or predictions about the grouping of anindividual data point.

Xgboost: XGBoost is a robust machine-learning algorithm that can help you understand your data and make better decisions. XGBoost is an implementation of gradient-boosting decision trees. It has been used by data scientists and researchers worldwide to optimize their machine-learning models.

Gradient Boosting: Gradient Boosting is a popularboosting algorithm in machine learning used for classification and regression tasks. Boosting is one kind of

ensemble Learning method which trains the model sequentially and each new model tries to correct theprevious model. It combines several weak learners into strong learners.

Random Forest: Random forest is a commonly-used machine learning algorithm trademarked by Leo Breiman and Adele Cutler, which combines the output of multiple decision trees to reach a single result. Its ease of use and flexibility have fueled its

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.134160

adoption, as it handles both classification and regression problems.

V. EXPERIMENTAL RESULTS

Accuracy: The accuracy of a test is its ability to differentiate the patient and healthy cases correctly. To estimate the accuracy of a test, we should calculate the proportion of true positive and true negative in all evaluated cases. Mathematically, this can be stated as:

Accuracy = TP + TN TP + TN + FP + FN.

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

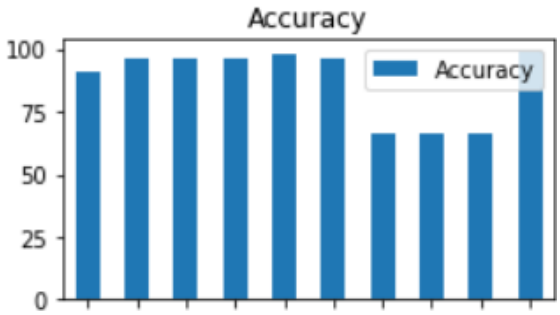
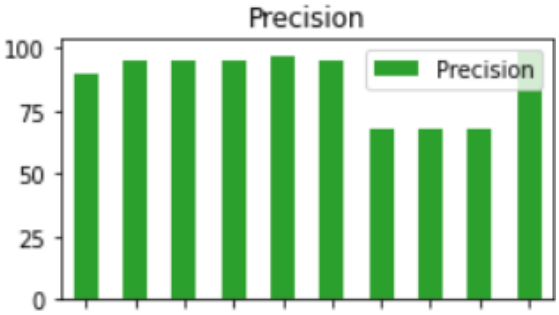


Fig 4.ACCURACY COMPARISON GRAPH

Precision: Precision evaluates the fraction of correctly classified instances or samples among the ones classified as positives. Thus, the formula to calculate the precision is given by:

Precision = True positives/ (True positives + Falsepositives) = TP/(TP + FP)

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$



$$\textbf{F1 Score} = \frac{2}{\left(\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}\right)}$$

$$\textbf{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.134160

Fig 5.PRECISION COMPARISON GRAPH

Recall: Recall is a metric in machine learning that measures the ability of a model to identify all relevantinstances of a particular class. It is the ratio of correctly predicted positive observations to the total actual positives, providing insights into a model's completeness in capturing instances of a given class.

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

Fig 6.RECALL COMPARISON GRAPH

F1-Score: F1 score is a machine learning evaluation metric that measures a model's accuracy. It combinesthe precision and recall scores of a model. The accuracy metric computes how many times a model made a correct prediction across the entire dataset.

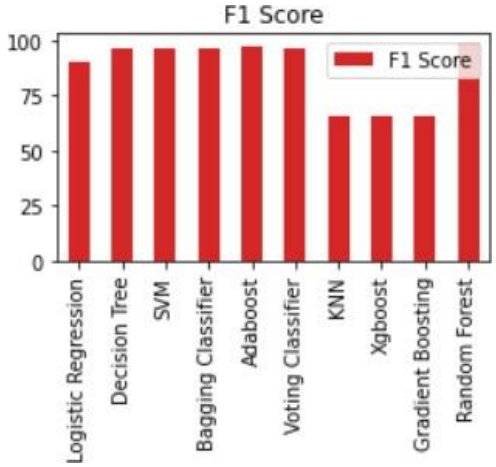


Fig 7. F1 SCORE COMPARISON GRAPH

	Accuracy	Recall	Precision	F1 Score
Logistic Regression	91.25	90.796703	90.162272	90.461591
Decision Tree	96.25	96.291209	95.571332	95.912110
SVM	96.25	96.291209	95.571332	95.912110
Bagging Classifier	96.25	96.291209	95.571332	95.912110
Adaboost	97.50	97.252747	97.252747	97.252747
Voting Classifier	96.25	96.291209	95.571332	95.912110
KNN	66.25	69.093407	67.473287	65.817376
Xgboost	66.25	69.093407	67.473287	65.817376
Gradient Boosting	66.25	69.093407	67.473287	65.817376
Random Forest	98.75	98.214286	99.056604	98.614719

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

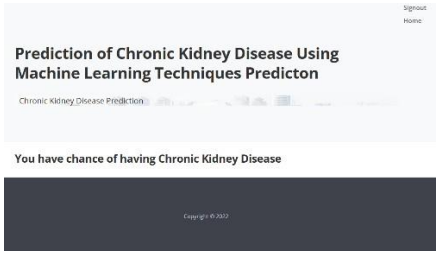
International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.134160

Fig 8. PERFORMANCE EVALUATION

Fig 9 HOME PAGE



ARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940
rch in Computer and Communication Engineering
& Refereed journal ⁜ Vol. 13, Issue 4, April 2024
3/IJARCCE.2024.134160

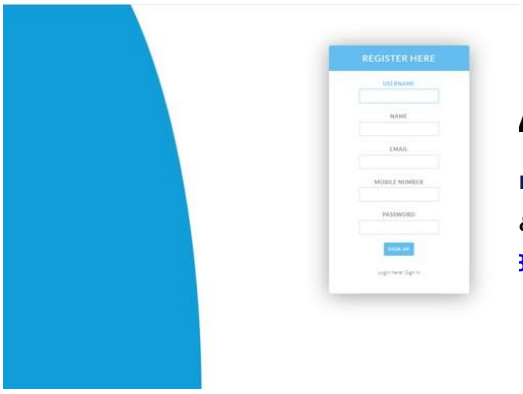


Fig 10 sign up

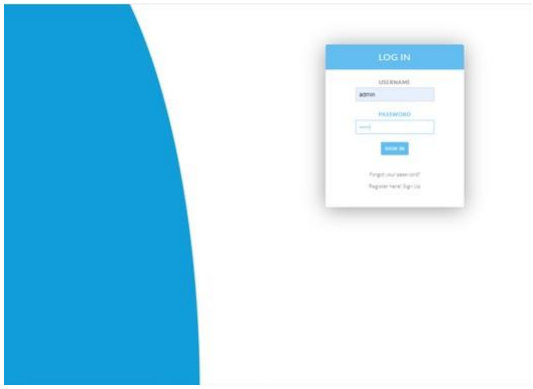


Fig 11 sign in

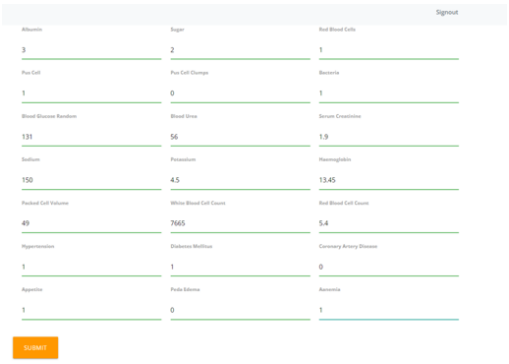


Fig 12 Data entry

Fig 13 Final output

5. CONCLUSION

In conclusion, the application of various machine learning algorithms and data mining techniques in predicting the progression of chronic kidney disease (CKD) has yielded promising results. Among the tested algorithms, Random Forest emerged as particularly effective, achieving an impressive accuracy of 98.75%. Leveraging a dataset comprising 400 instances with 21 distinct characteristics sourced from the UCI Repository, the Random Forest algorithm showcased its ability to discern patterns and make accurate predictions regarding CKD progression. This high level of accuracy underscores the potential of machine learning in healthcare, specifically in early identification andmonitoring of chronic conditions. Theensemble nature of Random Forest,combining multiple decision trees, likely contributed to its robustperformance. The findings highlightthe significance of incorporatingadvanced analytics in healthcaredecision-making processes to enhance diagnostic accuracy, facilitate timely interventions, and ultimately improve patient outcomes in the context of chronic kidney disease.

6. FUTURE SCOPE

In conclusion, the application of various machine learning algorithms and data mining techniques in predicting the progression of chronic kidney disease (CKD) has yielded promising results. Among the tested algorithms, Random Forest emerged as particularly effective, achieving an impressiveaccuracy of 98.75%. Leveraging a dataset comprising 400 instances with 21 distinct characteristics sourced from the UCI Repository, the Random Forest algorithm

DOI: 10.17148/IJARCCE.2024.134160

showcased its ability to discern patterns and make accurate predictions regarding CKD progression. This high level of accuracy underscores the potential of machine learningin healthcare, specifically in early identification and monitoring of chronic conditions. The ensemble nature of Random Forest, combining multiple decision trees, likely contributed to its robust performance. The findings highlight the significance of incorporating advanced analytics in healthcare decision-making processes to enhance diagnostic accuracy, facilitate timely interventions, and ultimately improve patient outcomes in the context of chronic kidneydisease.

6.REFERNCES

[1] Q.-L. Zhang and D. Rothenbacher, ‘‘Prevalence of chronic kidney disease in population-based studies: Systematic review,’’ BMC Public Health, vol. 8, no. 1, p. 117, Dec. 2008.

[2] W. M. McClellan, D. G. Warnock, S. Judd, P. Muntner, R. Kewalramani, M. Cushman, L. A. McClure, B. B. Newsome, and G. Howard, ‘‘Albuminuria and racial disparities in the risk for ESRD,’’ J. Amer. Soc. Nephrol., vol. 22, no. 9, pp.
1721–1728, Aug. 2011.

[3] M. K. Haroun, ‘‘Risk factors for chronic kidney disease: A prospective study of 23,534 men and women in Washington County, Maryland,’’ J. Amer. Soc. Nephrol., vol. 14, no. 11, pp. 2934–2941, Nov. 2003.

[4] W. D. Souza, L. C. D. Abreu, L. G. D. SilvaI, and I. M.
P. Bezerra, ‘‘Incidence of chronic kidney disease hospitalisations and mortality in Espírito Santo between 1996 to 2017,’’ Wisit Cheungpasitporn, Univ. Mississippi Medical Center, Rochester, MN, USA, Tech. Rep., 2019, doi:10.1371/journal.pone.0224889.

[5] W. Mula-Abed, K. A. Rasadi, and D. Al-Riyami, ‘‘Estimated glomerular filtration rate (eGFR): A serum creatinine-based test for the detection of chronic kidney disease and its impact on clinical practice,’’ Oman Med. J., vol. 27, no. 4, pp. 339–340, 2012.

[6] A. S. Levey, D. Cattran, A. Frie

dman, W. G. Miller, J. Sedor, K. Tuttle, B. Kasiske, and
T. Hostetter, ‘‘Proteinuria as a surrogate outcome in CKD: Report of a scientific workshop sponsored by the national kidney foundation and the US food and drug administration,’’ Amer. J. Kidney Diseases, vol. 54, no. 2, pp. 205–226, Aug. 2009.

[7] S. Gerogianni, ‘‘Concerns of patients on dialysis: Aresearch study,’’ Health Sci. J., vol. 8, no. 4, pp. 423– 437, 2014.

[8] J. R. Chapman, ‘‘What are the key challenges we face in kidney transplantation today?’’ Transplantation Res., vol. 2, no. S1, pp. 1–7, Nov. 2013.

[9] T. Xiuyi and G. Yuxia, ‘‘Research on application of machine learning in data mining,’’ in Proc. IOP Conf., Mater. Sci. Eng., 2018, doi: 10.1088/1757-899X/392/6/06220.

[10] B. Zupan, A. J. Halter, and M. Bohanec,
‘‘Qualitative model approach to computer assistedeasoning in physiology,’’ in Proc. Intell. Data Anal. Med. Pharmacol. (IDAMAP), Brighton, U.K., 2018, pp.1–7.

[11] A. Dhillon and A. Singh, ‘‘Machine learning in healthcare data analysis: A survey,’’ J. Biol. Today’s World, vol. 8, no. 2, pp. 1–10, Jan. 2018.

[12] M. Ghassemi, T. Naumann, P. Schulam, A. L.Beam, I. Y. Chen, and R. Ranganath, ‘‘A review of challenges and opportunities in machine learning for health,’’ in Proc. AMIA Joint Summits Transl. Sci., 2020, p. 191.

[13] J. Qin, L. Chen, Y. Liu, C. Liu, C. Feng, and B. Chen, ‘‘A machine learning methodology for diagnosing chronic kidney disease,’’ IEEE Access, vol. 8, pp. 20991–21002, 2020.

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.134160

[14] G. R. Vasquez-Morales, S. M. Martinez- Monterrubio, P. Moreno-Ger, and J. A. Recio-Garcia, ‘‘Explainable prediction of chronic renal disease in thecolombian population using neural networks and case-based reasoning,’’ IEEE Access, vol. 7, pp.152900–152910, 2019.

[15] Z. Chen, X. Zhang, and Z. Zhang, ‘‘Clinical risk assessment of patients with chronic kidney disease byusing clinical data and multivariate models,’’ Int. Urol.Nephrol., vol. 48, no. 12, pp. 2069–2075, Jun. 2016.

[16] Y. Amirgaliyev, S. Shamiluulu, and A. Serek, ‘‘Analysis of chronic kidney disease dataset by applying machine learning methods,’’ in Proc. IEEE 12th Int. Conf. Appl. Inf. Commun. Technol. (AICT), Oct. 2018, pp. 1–4.

[17] K. R. A. Padmanaban and G. Parthiban, ‘‘Applying machine learning techniques for predicting the risk of chronic kidney disease,’’ Indian J. Sci. Technol., vol. 9,no. 29, Aug. 2016.

[18] L. Kilvia De Almeida, L. Lessa, A. Peixoto, R. Gomes, and J. Celestino, ‘‘Kidney failure detection using machine learning techniques,’’ in Proc. 8th Int. Workshop ADVANCEs ICT Infrastructures Services, 2020, pp. 1–8.

[19] W. Gunarathne, K. D. M Perera, and K. A. D. C. P Kahandawaarachchi, ‘‘Performance evaluation onmachine learning classification techniques for disease classification and forecasting through data analytics for chronic kidney disease (CKD),’’ in Proc. IEEE 17th Int. Conf. Bioinf. Bioeng. (BIBE), Oct. 2017, pp. 291– 296.

[20] H. Polat, H. D. Mehr, and A. Cetin, ‘‘Huseyinpolat1 & homay danaei mehr1 & aydin cetin,’’ J. Med.Syst., vol. 41, no. 4, pp. 1–11, Apr. 2017, doi: 10.1007/s10916-017-0703-x. 17332 VOLUME 9, 2021
P. Chittora et al.: Prediction of Chronic Kidney Disease