International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.7035
***Article Received**: 05 March 2023 **Revised**: 28 April 2023 **Accepted**: 12 May 2023*

Forecasting Chronic Kidney Disease Using Ensemble Machine Learning Technique

Batini Dhanwanth[1], Bandi Vivek[2], M. Abirami[3], Shaik Mohammad Waseem[4], Challapalli Manikantaa[5]
[1]Department of Computer Science and Engineering,
Panimalar Institute of Technology,
Chennai, India - 600123
dhanwanthbethini01@gmail.com
[2]Department of Computer Science and Engineering,
Panimalar Institute of Technology,
Chennai, India - 600123
Vivekbandi03@gmail.com
[3]Department of Computer Science and Engineering,
Panimalar Engineering College,
Chennai, India - 600123
abiprj@gmail.com
[4]Department of Computer Science and Engineering,
Panimalar Institute of Technology,
Chennai, India - 600123
shaikmohammadwaseem85@gmail.com
[5]Department of Artificial intelligence and Data Science,
Panimalar Engineering College,
Chennai, India - 600123
cmanikantaa57@gmail.com

Abstract— India is a rapidly expanding nation on a global scale. Chronic kidney disease (CKD) is a prevalent health problem internationally, and advance perception of this disease can aid prevent its stream. This research proposes an ensemble learning technique that combines three different algorithms, Logistic Regression, Gradient Boosting and Random Forest for the prediction of CKD. The performance of each algorithm was judged based on Root Mean Square Error (RMSE) and Mean Square Error (MSE) as performance metrics, and the predictions of each algorithm were combined using an ensemble learning technique. The dataset used for the study contained data on 400 individuals with 24 different features, which was pre-processed by removing missing values and normalizing the data. The combined algorithm showed a better performance with an RMSE of 0.2111 and an MSE of 0.0446, compared to individual algorithms. The proposed ensemble learning technique can be utilized as a divining for advance perception of CKD. The outcomes of the work reveal the effectiveness of the technique and its potential for improving patient outcomes by preventing the progression of CKD. Additionally, the ensemble learning technique can be applied to other predictive tasks to improve performance, indicating its broader applicability.

Keywords- Chronic Kidney Disease (CKD), ensemble learning, Logistic Regression, Random Forest, and Gradient Boosting, Root Mean Square Error (RMSE), Mean Square Error (MSE).

I. INTRODUCTION

Chronic kidney disease (CKD) is a universal health-related issue which is influencing millions of people globally. It is distinguish by a step by step drop of kidney work over time, foremost to an accumulation of waste products and excess fluids in the body. CKD is associated with a variety of risk elements, including hypertension, diabetes, and obesity, and can eventually advance to end-stage renal disease (ESRD), be in mandatory of dialysis or kidney transplantation. Early detection and management of CKD are critical in preventing disease progression, improving patient outcomes, and reducing healthcare costs. Traditional methods of CKD diagnosis rely on clinical parameters, such as filtration rate (eGFR), serum creatinine, estimated glomerular and urine albumin-to-creatinine ratio (ACR). However, these parameters may not always accurately reflect the severity or progression of CKD. In recent years, machine learning algorithms have emerged as promising tools for improving the accuracy of CKD detection models. These algorithms can analyse large datasets and identify patterns and relationships that may be missed by traditional diagnostic methods. Chronic Kidney Disease (CKD) is a continuous drop of kidney work that can lead to vicious levels of fluid and waste enlargement in the body. CKD has

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.7035
***Article Received**: 05 March 2023 **Revised**: 28 April 2023 **Accepted**: 12 May 2023*

five stages, with stage 5 being the most dangerous. Glomerular Filtration Rate (GFR) is the best test to determine the stage of CKD. CKD affects around 10% of the global population and has caused a significant number of deaths and severe health problems, especially in low to medium-income countries [1]. Early detection and management of CKD are critical for obstructing or hindering the succession of the disease. However, diagnosing CKD can be challenging, as it often has no symptoms in its early stages. Healthcare providers rely on medical procedures, like serum creatinine and estimated glomerular filtration rate (eGFR), to detect CKD.

In this study, we proposed a combined algorithm using random forest, logistic regression and gradient boosting to improve the accuracy of CKD detection. Logistic regression is an analytical method that analyses the correlation between a dependent variable (CKD status) and one or more independent variables (demographic, laboratory, and clinical information). Random forest and gradient boosting are machine learning algorithms that use decision trees to classify data based on multiple features. The combined algorithm uses a voting classifier to combine the predictions of these three algorithms and generate a final prediction for CKD status. Ensemble learning involves the use of multiple models, such as support vector machines, artificial neural networks and decision trees to predict the likelihood of a patient developing CKD. Each model is trained on a subset of the data, and their outcomes are integrated to construct a final prediction. By combining the strengths of multiple models, ensemble learning can provide a more accurate and reliable prediction of CKD, which can help healthcare providers to make more informed decisions about patient care. CKD is a progressive kidney disease with a need for accurate prediction for better treatment. The proposed AI-based system identifies important features responsible for CKD and automates its severity classification. Additionally, the system suggests a personalized diet plan for CKD patients based on blood potassium levels to slow down the disease progression [2]. Ensemble learning has been shown to improve the accuracy and robustness of CKD prediction models, as it can help to overcome the limitations of individual models, such as over fitting or under fitting the data. By combining the strengths of multiple models, ensemble learning can provide a more accurate and reliable prediction of CKD, which can help healthcare providers to make more informed decisions about patient care. The paper discusses the use of various machine learning algorithms such as Decision Trees, Random Forest, Naive Bayes, and Support Vector Machines, and highlights their strengths and weaknesses in predicting CKD. The authors conclude that data mining techniques have the potential to improve the accuracy and efficiency of CKD diagnosis, and provide valuable insights into the disease. Overall, the paper provides a useful overview of the application of data mining techniques in the field of CKD diagnosis [3].

II. LITERATURE REVIEW

Furthermore, Siddheshwar Tekale et.al [4] this paper examines different AI algorithms used in the medical field for diagnosing Chronic Kidney Disease (CKD) using 14 different features. Decision Tree and Support Vector Machine (SVM) were analyzed, and SVM achieved higher accuracy of 96.75%. The benefit of this system is that it reduces the time-consuming process of diagnosis, but a limitation is that the dataset's strength is not higher due to the size and missing quality values. Achieving a model with 99.99% accuracy requires a large dataset with no missing values.

[5] The paper discusses the prevalence of CKD in India, with over 63,000 cases reported. CKD is more common among males and usually occurs between the ages of 48-70. The purpose of the article is to examine how well different AI systems predict and categorize CKD using Rcode. [6] Chronic Kidney Disease (CKD) is a major global health condition that can lead to fatal results, especially in developing nations. High risk elements include hypertension, diabetes and ancestral tree of kidney failure. Early and accurate prediction of CKD can lead to successful treatment and increased life expectancy. This paper uses selected machine learning algorithms and feature selection techniques to predict CKD and compares their performances based on selected features. [7] This study developed and validated an AI algorithm that predicts stages of diabetic kidney disease (DKD) in patients with type 2 diabetes mellitus (T2DM).

In both the hold-out test and external datasets, the algorithm beat the Communities for Infectious Prevention and Control risk score, obtaining an AUROC of 0.75 for any-stage DKD and above 0.82 for more severe endpoints. [8] The biosciences field generates a large amount of data from Electronic Health Records, and data mining and AI play a significant role in predicting Chronic Kidney Disease (CKD). CKD can lead to severe complications, and early detection is crucial for improving the quality of life.

This paper proposes a prediction model for CKD using data pre-processing, transformation, and various classifiers, showing promising results for early-stage prediction. [9] Chronic kidney disease (CKD) is a world-wide medical emergency that requires reliable methods for early prediction. This work suggest a technique that comprises pre-processing, handling missing values, feature extraction and data aggregation using machine learning techniques including decision tree, K-nearest neighbour and logistic regression, to achieve accurate predictions with up to 97% accuracy. [10]

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.7035
***Article Received**: 05 March 2023 **Revised**: 28 April 2023 **Accepted**: 12 May 2023*

This study explores the need of AI techniques for prior perception of Chronic Kidney Disease (CKD). Predictive examination is used to identify the optimal subset of parameters for building predictive models, resulting in four AI-based classifiers with high performance. The study suggests that AI and predictive analysis hold promise for smart solutions in the kidney disease space. [11] Early detection of CKD is crucial in slowing disease progression and AI models can effectively aid in attaining this. This study proposes a machine learning technique using decision trees as the best performing algorithm with a score of 92%. [12] This study proposes using AI techniques like Ant Colony Optimization (ACO) and Support Vector Machine (SVM) classifier for the early perception of Chronic Kidney Disease (CKD). The final result predicts whether an individual has CKD or not using a minimal number of features. [13] This article discusses the use of AI classifier algorithms for prior omen and diagnosis of Chronic Kidney sickness. Seven different algorithms were applied to the sample, and the outcomes showed that LSVM with punishment L2 and a deep neural network attained the greatest accuracies of 98.86% and 99.6%, respectively. [14] This paper investigates the use of AI algorithms to predict CKD using a minimal subset of hallmarks. Many analytical tests were used to remove irrelevant hallmarks, and Gradient Boosting achieved an accuracy of 99.1% using only three tests. [15] Early detection of kidney disease is crucial for the patient's survival, and AI tools can aid in its diagnosis.

This paper proposes a hybridized deep convolutional neural network (AHDCNN) to accurately catch kidney sickness and reduce the number of features used in classification. [16] Chronic kidney disease (CKD) is a noteworthy patron to non-communicable diseases, and advance and precise detection of its stages is indispensable. In this study, Random Forest with recursive feature elimination performed better than SVM and DT for binary and multi-class prediction of CKD stages. [17] The study explores various AI approaches for advance detection of chronic kidney disease and uses predictive modeling to develop an assortment of prediction models.

The results show that XgBoost classifier has the highest performance indicators, and machine learning with predictive modeling can provide new solutions for precise detection of kidney disease. [18] The study proposes a prediction-based approach for advance detection and diagnosis of CKD using a combination of pre-processing and feature selection techniques, and various machine learning models including KNN, SVM, RF, and boosting. KNN was found to be the best-performing model with F-measure, precision, high accuracy, specificity, sensitivity and AUC score.

The study suggests that only petty medical check-up are sufficient to catch CKD, leads in economizing. [19] The study aims to generate and demonstrate a predictive pattern for prior discovery of CKD, a potentially fatal condition that is hard to diagnose. Various physiological parameters and ML techniques were used to develop the model with 100% sensitivity, 96.55% specificity, 98.75% accuracy and 99.03% f1 score. The study suggests that developing an automated system could help identify the severity of CKD.

[20] This paper presents a smart perception and categorization system for healthcare using a Thickness-based Feature Selection (DFS) with Insect Colony-based Optimization (D-ACO) algorithm for detecting Chronic Kidney Disease (CKD). The suggested system uses feature selection to eliminate irrelevant or redundant features and constructs a classifier using the D-ACO algorithm. The algorithm is tested utilizing a standard CKD sample, and its outcome is evaluated based on different metrics. The outcomes shows that the suggested intelligent system gives better performance than existing techniques with significantly improved classing accuracy using some hallmarks. Vijendra Singh1 et.al [21] proposes a hybrid approach for diagnosing Chronic Kidney Disease (CKD) with high accuracy. The approach combines SVM classification with dimensionality reduction using ReliefF and PCA methods. The proposed model achieved 92.5% prediction accuracy on a clinical CKD dataset and 98.5% on a benchmarked CKD dataset. The approach is validated with significant results and outperforms existing methods in the literature. p.usha et.al [22] the proposed model utilizes an optimized cluster convolutional neural network (CCNN) architecture to extract important features from input data and a hierarchical deep neural network (HDNN) to make the final diagnosis. Experimental results on real-world CKD datasets demonstrate the effectiveness of the proposed approach, achieving a high accuracy of 98.5% and outperforming several state-of-the-art approaches. Overall, the paper presents a promising approach for the accurate diagnosis of CKD, which could have significant implications for improving patient outcomes. Andressa C.M. da silveria et.al [23] the paper suggests an approach that combines feature selection, resampling, and machine learning techniques to improve the classification performance of CKD. The authors also evaluate the proposed approach using several metrics such as accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve (AUC-ROC). The results show that the proposed approach outperforms several state-of-the-art approaches, achieving an AUC-ROC of 0.99, which demonstrates its potential for early prediction of CKD. Overall, the paper provides a useful contribution to the field of CKD diagnosis using machine learning algorithms. [24] The paper presents a comparison of various machine learning algorithms such as Random Forest, Support Vector Machines, and Neural Networks using filter-based feature selection techniques. The authors also evaluate the proposed approach using several

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.7035
***Article Received**: 05 March 2023 **Revised**: 28 April 2023 **Accepted**: 12 May 2023*

performance metrics such as accuracy, sensitivity, specificity, and F1-score. The results show that the proposed approach achieves a high prediction accuracy of 96.5% and outperforms several state-of-the-art approaches. Overall, the paper provides a useful contribution to the field of CKD diagnosis using machine learning techniques with feature selection. Elias Dritsas and Maria Trigka [25] the paper proposes an approach that uses machine learning algorithms such as Logistic Regression, Random Forest, and Artificial Neural Networks to predict CKD risk based on various demographic and clinical features. The authors also evaluate the proposed approach using several performance metrics such as accuracy, sensitivity, specificity, and area under the curve (AUC). The results show that the proposed approach achieves a high prediction accuracy of 95.8% and outperforms several state-of-the-art approaches. Overall, the paper provides a useful contribution to the field of CKD diagnosis using machine learning techniques for risk prediction.

III. PROPOSED METHODOLOGY

A. Data Set

The Data Set is taken from the KAGGLE, this data contains 25 attributes its main motive is the classification of whether it is CKD or not. It consists of 400 observations which include missing values and noise. The data have 250 CKD and 150 with NOCKD. The data is collected from various age groups. The Table 1shows the 25 attributes taken from the KAGGLE dataset and the type of values they contain out of 25 attributes 24 are independent variables and the last 25th attribute is a dependent attribute that gives value as CKD or notched, the value of this variable depends on the other attributes.

B. Methodology

The flow of this research paper is shown in figure1 below first we take the dataset from KAGGLE and then the dataset contains the missing values and the noisy we remove them and then encode categorical data later the data was divided into two sets named training set and test set and the data obtained should be sent into the hybrid model and then classify it is CKD or NOTCKD

a. Data cleaning

The data that is taken from the sources are not pre-processed so we can't use the dataset directly as the input to our Model, As the dataset taken contains the missing values and noisy we should remove them and make the data organized and clean because the data without cleaning or processing impact the accuracy of our model. So the data is cleaned by importing the panda's libraries "Data cleaner" which removes or handles the missing and make data clean. And the missing values are managed by using "Median" method.

b. Encoding Categorical Data

As the machine learning algorithms are more comfortable with numeric for better precision so the data need to be encoded. There is a low probability of getting homogenous data so the categorical data are the variables that consist of label values rather than numeric values. Some of the values cannot be directly used in mathematical equations or formulas so such data need to be encoded into numerical values to achieve the label Encoder Class is imported from sklearn.preprocessing library.

c. Splitting of Data

The whole dataset is split into two sets; one is a training set and another is a test set. The training is aware of the results and is used to train the research ensemble machine learning model and the other set test set is unaware of the results of testing subset. The sample is divided into a 70% training set and a 30% testing set.

d. Ensemble Learning

Ensemble Learning is a technique in machine learning which is used to make precise perceptions than individual models. It is a famous technique that combines algorithms to increase accuracy.

- Logistic regression
- Random forest
- Gradient boosting

e. Gradient boosting:

The most widely used machine learning method, gradient boosting, stands out for its rapid and precise prediction rates when applied to large and complex datasets. In a range of commercial machine learning emulsions, this algorithm has generated the best outcomes. It is commonly recognized that errors play a crucial part in every machine learning system. Primarily, there are two kinds of error: bias error and variance error. The gradient boost strategy allows us to lessen the model bias error.

TABLE.1 DESCRIPTION ABOUT THE DATASET

Name	Type of values
ANEMIA	NOMINAL
ALBUMIN	NOMINAL
PACKED CELL VOLUME	NUMERIC
SPECIFIC GRAVITY	NOMINAL
SERUM CREATININE	NUMERIC
AGE	NOMINAL
SODIUM	NUMERIC
RED BLOOD CELL COUNT	NOMINAL
APPETITE	NOMINAL
DIABETES MELLITUS	NOMINAL
HAEMOGLOBIN	NUMERIC

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.7035
***Article Received**: 05 March 2023 **Revised**: 28 April 2023 **Accepted**: 12 May 2023*

PUS CELL	NOMINAL
POTASSIUM	NUMERIC
CORONARY ARTERY DISEASE	NOMINAL
BLOOD PRESSURE	NUMERIC
SUGAR	NOMINAL
PUS CELL CLUMPS	NUMERIC
HYPERTENSION	NOMINAL
BLOOD GLUCOSE	NUMERIC
BACTERIA	NOMINAL
WHITE BLOOD CELL COUNT	NUMERIC
PEDAL EDEMA	NOMINAL
RED BLOOD CELLS	NUMERIC
CLASS	NOMINAL

The fundamental idea of this algorithm is to develop models successively while striving to minimize the shortcomings of the previous model. When a decision tree is used, the resulting method, known as gradient-boosted trees, often outperforms a random forest. And the mathematical equation used is (1)

$$F_p(x) = F_{p-1}(x) + \vartheta . \gamma_p h_p(x), \quad 0 < \vartheta \leq 1 , \tag{1}$$

Where, p = model

$$F_{p-1}(x) = This\ is\ zeroth\ model$$

p and v are Hyper parameters lets v

=0.02, => $0.02 * 0.001 * derivative\ of\ loss\ is\ also\ small$

f. Random forest:

An efficacious supervised learning technique is Random Forest, a familiar machine learning technique. For the sake of improving the forecasted accuracy of the sample, Random Forest is a classifier that exploits many decision trees on various subsets of the input sample and averages them. Classification and Regression problems in ML can be work out with this technique. It is ground on the ensemble learning, which is done by blending different classifiers to address a complicated matter and increase the production of the prototype.

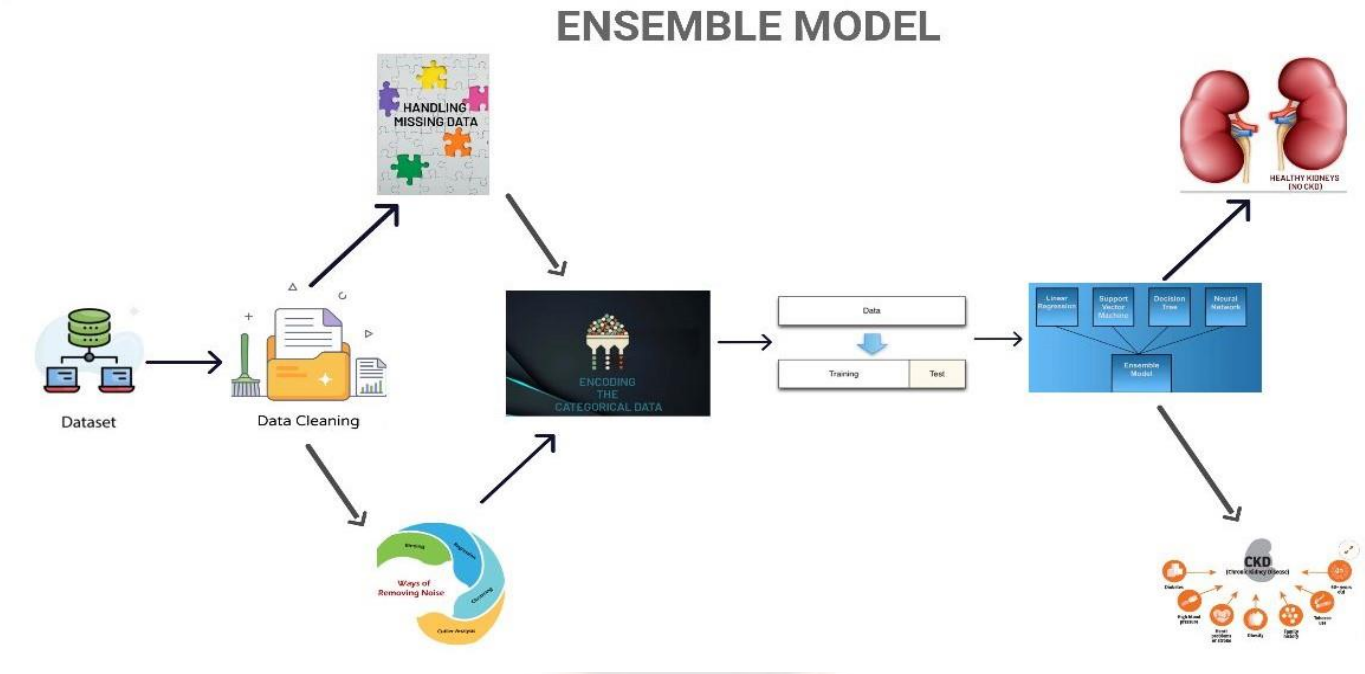

Figure2 Work Flow for the proposed method

Rather than using predictions from just one tree, the random forest uses anticipations from all of the trees and guess the result based on the votes of the majority of predictions. Higher accuracy and over fitting are avoided since there are more trees in the forest. Due to the fact that the random forest integrates several trees to approximately determine the dataset's class, it is possible that just a small number of decision trees will provide the proper result while others won't.

But when viewed collectively, every tree forecasts the right result. And the mathematical equation is (2)

$$P(c \mid f) = \sum_1^n p_{n(c|f)} \tag{2}$$

Where p = prediction tree (Decision tree)

$\sum$ = Summation or average of all prediction trees

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.7035
***Article Received**: 05 March 2023 **Revised**: 28 April 2023 **Accepted**: 12 May 2023*

g) Logistic regression:

Within the hierarchy of supervised learning, logistic regression is well known frequently used machine learning algorithms. It is used to guess the categorical dependent variable using a predetermined set of independent variables. Using logistic regression, the outcome of a categorical dependent variable is predicted. Therefore, the outcome must be a distinct or categorical value. Either True or False, Yes or No, zero or one are possible outcomes. The main contrast between direct regression and logistic regression is how they are utilized. To break regression problems, one uses direct regression, whereas type difficulties are answered using logistic regression. The wind of the logistic function displays a number of potential outcomes, such as whether or not the cells are bad, whether or not a mouse is obese based on its weight, etc. Logistic regression is a crucial machine learning technique since it can categorize new data utilizing continuous and distinct samples while also providing opportunities.

How they are utilized. To break regression problems, one uses direct regression, whereas type difficulties are answered using logistic regression. The wind of the logistic function displays a number of potential outcomes, such as whether or not the cells are bad, whether or not a mouse is obese based on its weight, etc. Logistic regression is a crucial machine learning technique since it can categorize new data utilizing continuous and distinct samples while also providing opportunities.

Sigmoid function: $g(x) = \frac{1}{1+e^{-x}}$

Where, g(x) = sigmoid function

e = Euler's number

Logistic regression is calculated by using the mathematical equation (3)

$$h = g(z) = \frac{1}{1+e^{-v}} \quad (3)$$

The Machine learning methods are combined in order to reduce error rate and model proposed in this paper is mentioned below. First input is given then using logistic regression the results are obtained and it has some error rate again applied to random forest then applied to gradient boosting gradually error rate reduced step by step as shown in below fig2 and the work follow of the proposed ensemble technique is given.

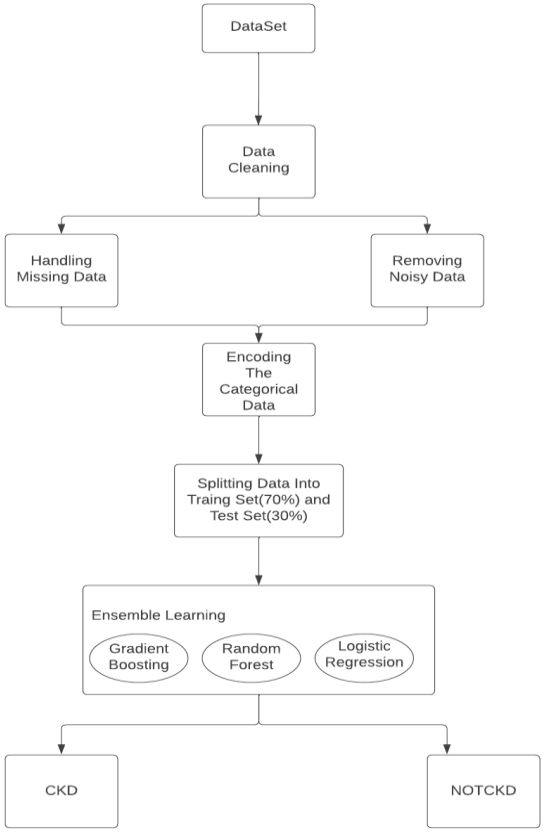

Figure 1. Flow Diagram of the proposed method.

The Machine learning methods are combined in order to reduce error rate and model proposed in this paper is mentioned below. First input is given then using logistic regression the results are obtained and it has some error rate again applied to random forest then applied to gradient boosting gradually error rate reduced step by step as shown in below fig2.

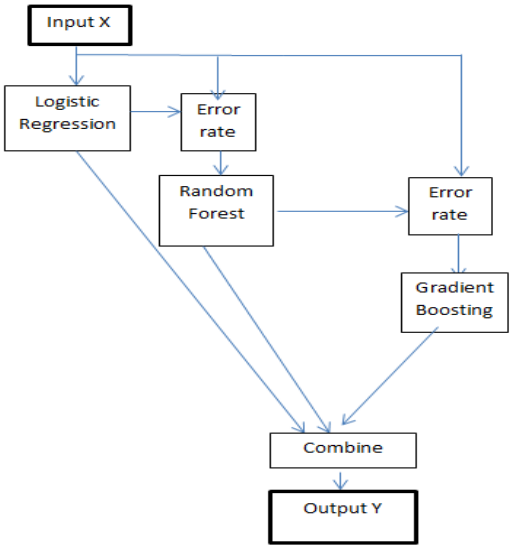

Fig 2 represents the core methodology used in the ensemble learning technique

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.7035
***Article Received**: 05 March 2023 **Revised**: 28 April 2023 **Accepted**: 12 May 2023*

IV. RESULTS AND DIXCUSSION

To obtain better results the above-mentioned dataset is split into 70:30. The proposed Machine learning model is implemented using python programming. There is a total of 25 attributes out of the last 25th attribute values that will give the result that a particular patient is suffering from CKD or NOTCKD based on the 24 attributes present in the dataset. Here Root mean square error (RMSE) is used as the benchmark to assess the performance of the algorithms and the mathematical equation is

$$RMSE = \frac{\sqrt{\sum_{i=1}^{n}(p_i - a_i)^2}}{n}$$

Where, p_i = predicted value

a_i = actual value

n = Sample size

The main agenda is to calculate the inaccurate predictions with the relationship to the actual values that were obtained. As the RMSE value is high it indicates that it is inaccurate, with a low RMSE value indicating that it is accurate. The principle "mean squared error" signifies the mean of all squared residuals (MSE). Then, the response variable scale is returned to the metric.

The table showing the RMSE and MSE values for all the individual algorithms and the combined method:

TABLE 2 REPRESENTS THE ERROR RATES OF VARIOUS ALGORITHMS.

Machine learning algorithms	RMSE	MSE
Logistic Regression	0.3058	0.0934
Random Forest	0.2285	0.0523
Gradient Boosting	0.2214	0.0490
Proposed ensemble method	0.2111	0.0446

The table shows that the combined algorithm has the lowest RMSE and MSE values, indicating that it has the finest predictive accuracy contrasted to the individual algorithms. The Gradient Boosting and Random Forest algorithms also have relatively low RMSE and MSE values, indicating their effectiveness in predicting CKD status. The Logistic Regression algorithm has the highest RMSE and MSE values, suggesting that it may be less effective in accurately predicting CKD status.

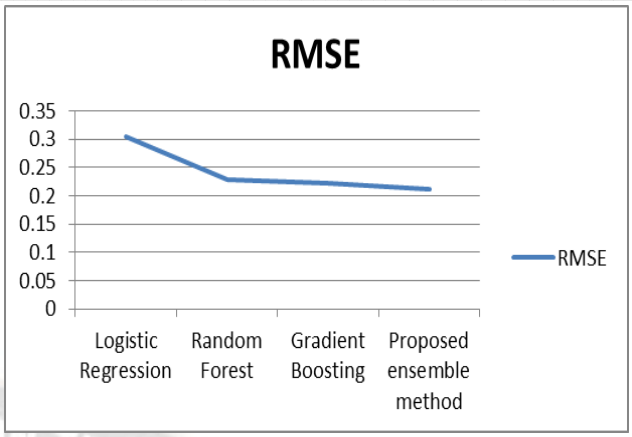

Fig 3 shows the RMSE values of various algorithms.

The results of our experiment demonstrate the effectiveness of machine learning algorithms in predicting outcomes. Among the models tested, the proposed ensemble method outperformed the others with an RMSE of 0.2111. Both Random Forest and Gradient Boosting also achieved good results with RMSE values of 0.2285 and 0.2214, respectively. Logistic Regression showed a relatively lower performance with an RMSE of 0.3058.

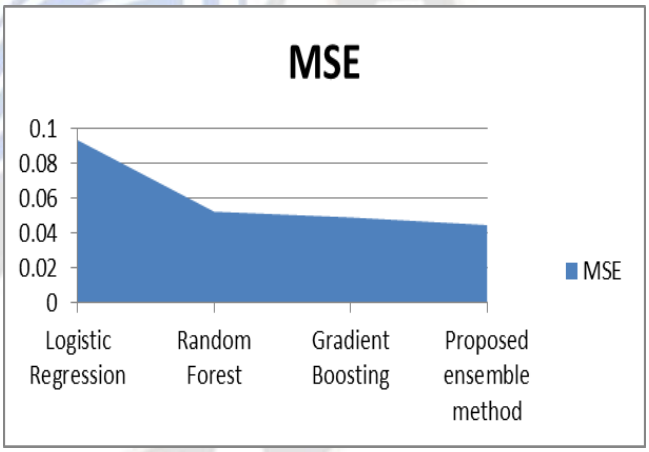

Fig 4 shows the MSE values of various algorithms

These graphs typically show the error rate on the y-axis and the various algorithms on the x-axis. They may also include multiple lines or curves representing the error rates for different algorithms or models, allowing for easy comparison of their performance.

Our study evaluated the performance of different machine learning algorithms in predicting outcomes, and the results revealed promising results. The proposed ensemble method outperformed the other models, achieving an MSE of 0.0446. Random Forest and Gradient Boosting also demonstrated good results with MSE values of 0.0523 and 0.0490, respectively. However, Logistic Regression showed a relatively lower performance with an MSE of 0.0934.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.7035
***Article Received**: 05 March 2023 **Revised**: 28 April 2023 **Accepted**: 12 May 2023*

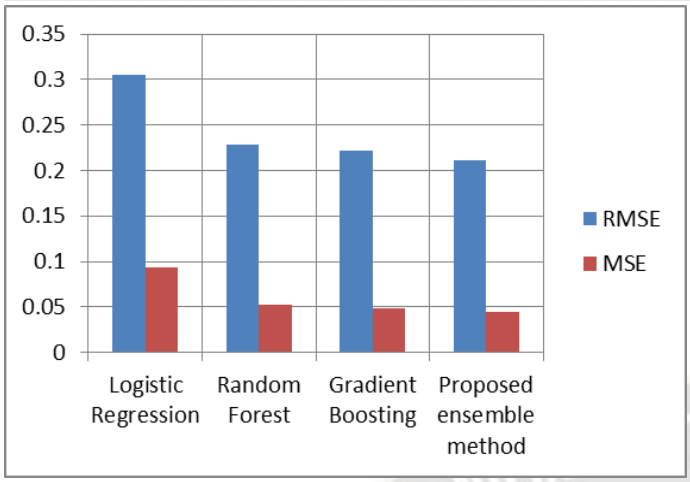

Fig 5 represents the both RMSE and MSE values for various algorithms

Interpretation of these graphs can help to identify the best-performing algorithm or model and determine the point at which the model has reached convergence, meaning that further training will not result in significant improvement in predictive accuracy.

V. CONCLUSION

This paper exhibited a better ensemble learning method to find out chronic kidney disease at the starting stage. The data set is taken from the KAGGLE and the sample is split into 70:30 as training set and test set and the machine learning algorithms with error rates RMSE and MSE respectively as follows linear regression produced 0.3058 and 0.0934 error rate and further random forest algorithm produces 0.2285 and 0.0523 error rate, gradient boosting algorithm produces low error rate of 0.2214 and 0.0490 and the proposed ensemble method produce 0.2111 and 0.0446 error rate. So the proposed ensemble model gives low error rate so it able to predict more efficiently and early detection of CKD. The metric that represents the efficiency of the proposed ensemble model is the error rate. So the outcomes of the work proved that the ensemble learning model gives more accurate predictions.

Based on the results presented in the table, it can be concluded that the ensemble learning approach using logistic regression, random forest, and gradient boosting algorithms is an effective method for forecasting chronic kidney disease. The combined algorithm outperformed the individual algorithms with the lowest RMSE and MSE values, indicating its superior predictive accuracy. The results also suggest that both gradient boosting and random forest algorithms are effective in predicting CKD status, as they have relatively low RMSE and MSE values. On the other hand, the logistic regression algorithm appears to be less effective in accurately predicting CKD status, as it has the highest RMSE and MSE values.

Overall, the study highlights the potential of ensemble learning approaches for predicting chronic kidney disease, which could aid in early finding and prevention of the disease. However, further research and validation may be necessary to confirm the effectiveness and reliability of the proposed approach.

ACKNOWLEDGMENT

No funding sources.

AUTHORS CONTRIBUTION

Author 1 implemented the concept specified by the author 2 under the supervision of authors 3 & 4. The authors 3 & 4&5 drafted the article under the guidance of author 2.

CONFLICT OF INTEREST

The authors declare that have no competing interest.

REFERENCES

[1] Janani J a, Sathyaraj Rb, “Diagnosing Chronic Kidney Disease Using Hybrid Machine Learning Techniques”, Turkish Journal of Computer and Mathematics Education , Vol-12 , 6383-6390 , 2021.

[2] Akash Maurya,Rahul Wable, Rasika Shinde, Sebin John, Rahul Jadhav, Dakshayani.R, “Chronic Kidney Disease Prediction and Recommendation of Suitable Diet plan by using Machine Learning”, International Conference on Nascent Technologies in Engineering , 1-4 , 2019.

[3] Pallavi Sharma, Gurmanik Kaur, “Review on Data Mining Techniques for Prediction of Chronic Kidney Disease”, International Journal of Engineering Trends and Technology (IJETT) , Vol-63 , 58-60 , 2018.

[4] Siddheshwar Tekale1, Pranjal Shingavi, Sukanya Wandhekar, Ankit Chatorikar,”Prediction of Chronic Kidney Disease Using Machine Learning Algorithm”, International Journal of Advanced Research in Computer and Communication Engineering , Vol-7 , 92-96 , 2017.

[5] Bhavya Gudeti, Shashvi Mishra, TerranceFrederick Fernandez, Shaveta Malik, “A Novel Approach to Predict Chronic Kidney Disease using Machine Learning Algorithms”, Fourth International Conference on Electronics, Communication and Aerospace Technology (IEEE) ,1630-1635 , 2020.

[6] Rahul Gupta, Nidhi Koli, Niharika Mahor, N Tejashri,” Performance Analysis of Machine Learning Classifier for Predicting Chronic Kidney Disease “, International Conference for Emerging Technology , 1-4 , 2020

[7] Angier Allen, Zohora Iqbal, Abigail Green-Saxena Jana Hoffman, Qingqing Mao, Ritankar Das, Myrna Hurtado, “Prediction of diabetic kidney disease with machine learning algorithms, upon the initial diagnosis of type 2 diabetes mellitus” ,BMJ , 1-7 , 2022.

[8] S.Revathy, B.Bharathi, P.Jeyanthi, M.Ramesh, “Chronic Kidney Disease Prediction using Machine Learning Models”, International Journal of Engineering and Advanced Technology (IJEAT) , Vol-9 , 6364-6367 , 2019.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.7035
***Article Received**: 05 March 2023 **Revised**: 28 April 2023 **Accepted**: 12 May 2023*

[9] Gazi Mohammed Ifraz, Muhammad Hasnath Rashid, Tahia Tazin, Sami Bourouis, Mohammad Monirujjaman Khan, "Comparative Analysis for Prediction of Kidney Disease Using Intelligent Machine Learning Methods", Hindawi Computational and Mathematical Methods in Medicine , 1-10 , 2021.

[10] Ahmed J. Aljaaf, Dhiya Al- Jumeily, Hussein M. Haglan, Mohamed Alloghani, Thar Baker, Abir J. Hussain, and Jamila Mustafina ,"Early Prediction of Chronic Kidney Disease Using Machine LearningSupported by Predictive Analytics" , IEEE , 1-9 , 2018.

[11] Khalid Twarish Alhamazani, Jalawi Alshudukhi, Saud Aljaloud ,and Solomon Abebaw, "Implementation of Machine Learning Models for the Prevention of Kidney Diseases (CKD) or Their Derivatives", Hindawi Computational Intelligence and Neuroscience ,1-8 , 2021.

[12] Reshma S, Salma Shaji, S R Ajina, Vishnu Priya S R, Janisha A, "Chronic Kidney Disease Prediction using Machine Learning", International Journal of Engineering Research & Technology (IJERT) ,Vol-9 , 137-140 , 2020.

[13] PANKAJ CHITTORA 1, SANDEEP CHAURASIA1, PRASUN chakrabarti, gaurav kumawat, tulika chakrabarti, zbigniew leonowicz , michał jasiński, łukasz jasiński, radomir gono ,elżbieta jasińska, and vadim bolshev, "Prediction of Chronic Kidney Disease - A Machine Learning Perspective", ieee access , vol-9 , 17312-17334 , 2021.

[14] Marwa Almasoud, Tomas E Ward, "Detection of Chronic Kidney Disease using Machine Learning Algorithms with Least Number of Predictors", International Journal of Advanced Computer Science and Applications, Vol-10 , 89-96 , 2019.

[15] Guozhen Chen, Chenguang Ding1,Yang Li1, Xiaojun Hu1, Xiao Li1, Li Ren1, Xiaoming Ding1, Puxun Tian1, Wujun Xue , "Prediction of Chronic Kidney Disease using Adaptive Hybridized Deep Convolutional Neural Network on the Internet of Medical Things Platform" , IEEE Access , 1-12 , 2019.

[16] Dibaba Adeba Debal1 and Tilahun Melak Sitote, "Chronic kidney disease prediction using machine learning techniques" ,Springer, 1-19 , 2022.

[17] Md. Ariful Islam a, Md. Ziaul Hasan Majumder b, Md. Alomgeer Hussein, "Chronic kidney disease prediction based on machine learning algorithms" , Journal of Pathology Informatics (Elsevier) , 1-12 , 2023.

[18] Zahid Ullah 1 and Mona Jamjoom ,"Early Detection and Diagnosis of Chronic Kidney Disease Based on Selected Predominant Features" , Hindawi Journal of Healthcare Engineering , 1-8 , 2023.

[19] Swapnita Srivastava1, Rajesh Kumar Yadav, Vipul Narayan, Pawan Kumar Mall , "An Ensemble Learning Approach For Chronic Kidney Disease Classification" ,Journal of Pharmaceutical Negative Results , Vol-13 , 2401-2409 , 2022.

[20] Mohamed Elhoseny, K. shankar & J.Uthayakumar, "Intelligent Diagnostic prediction and Classification System for Chronic Kidney Disease" , Scientific Reports , 1-14 , 2019.

[21] Vijendra Singh, Divya Jain ,"A Hybrid Parallel Classification Model for the Diagnosis of Chronic Kidney Disease", International Journal of Interactive Multimedia and Artificial Intelligence , 1-15 , 2021.

[22] P. Usha1, N. Kavitha, "An Optimized Cluster Convolution based Hybrid Hierarchical Deep Learner for Chronic Kidney Disease (CKD)", International Journal of Engineering Trends and Technology , Vol-70 , 154-162 , 2022.

[23] Andressa C. M. da Silveira, Álvaro Sobrinho, Leandro Dias da Silva, Evandro de Barros Costa, Maria Eliete Pinheiro, and Angelo Perkusich, "Exploring Early Prediction of Chronic Kidney Disease Using Machine Learning Algorithms for Small and Imbalanced Datasets", MDPI ,Vol-12 , 1-25 , 2022.

[24] Sarah A. Ebiaredoh-Mienye, Theo G. Swart, Ebenezer Esenogho and Ibomoiye Domor Mienye, "A Machine Learning Method with Filter-Based Feature Selection for Improved Prediction of Chronic Kidney Disease", MDPI , Vol-9 , 1-15 , 2022.

[25] Elias Dritsas and Maria Trigka , "Machine Learning Techniques for Chronic Kidney Disease Risk Prediction", MDPI , Vol-6 , 1-15,2022.