IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 11, Issue 5, May 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105158 |

Chronic Kidney Disease Prediction Using Machine Learning Techniques

Yashaswini A R[*1], Prathibha D K[*2], Ronish J Gowda[*3], Shivananda S Belagavi[*4], Siddharth Priyadarshi V[*5]

Assistant Professor, Dept. of Computer Science and Engineering, Maharaja Institute of Technology, Mysore, Karnataka, India[*1]

UG Student, Dept. of Computer Science and Engineering, Maharaja Institute of Technology, Mysore, Karnataka, India[*2,3,4,5]

ABSTRACT : Chronic Kidney Disease is one among the non-contagious illnesses that affect most of the individual in the world. The main factors of risk for the Chronic Kidney Disease are Diabetes, Heart Ailment, Hypertension. The Chronic Kidney Disease shows no symptoms in the early stages and most of the cases are diagnosed in the advanced stage. This leads to delayed treatment to the patient which may be fatal. Machine learning technique provides an efficient way in the prediction of Chronic Kidney Disease at the earliest stage. In this paper, four ensemble algorithms are used to diagnose the patient with Chronic Kidney Disease at the earlier stages. The machine learning models are evaluated based on different performance metrics including Accuracy, Sensitivity, Specificity. Based on the evaluation the XGBoost and Random Forest performed the best in terms of accuracy, precision, Sensitivity compared to Gradient Boosting and Bagging. The XGBoost and Random Forest also showed and Area Under the curve scores of 98%. The machine learning model proposed in this paper will provide an efficient way to prevent Chronic Kidney diseases by enabling the medical practitioners to diagnose the disease at an early stage.

KEYWORDS: Analysis, investigation, research (5-6 Keywords, Font-Cambria, Font Size – 10).

I. INTRODUCTION

Chronic Kidney Disease (CKD) is a major public health concern with rising prevalence. Kidney disease is when the kidneys are damaged and could not filter the blood properly. This damage could cause the wastes to build up in the body. There are five stages of CKD, the most serious one is stage 5 because, at this stage, the kidneys are unable to do most of their functions. It is difficult to pinpoint the CKD stage of each patient especially at the early-stages. It also causes a high possibility of death within a short period of time, a patient must be hospitalized and appropriately cured. The most common causes of kidney disease are diabetes and high blood pressure.

In Saudi Arabia, the number of people who need kidney replacement due to CKD is growing day by day. CKD is ranked the fourth leading cause of death in Saudi Arabia with the mortality rate of 5.44%. Moreover, the escalating prevalence of CKD poses challenges on professionals and the health care system. Approximately two billion Saudi Riyals are allocated for renal replacement therapy which is required for patients with advanced stages of CKD. Hence, the need for a preemptive diagnostic predictive solution that will facilitate early and easier accurate diagnosis of CKD cannot be overemphasized.

Machine learning is a field of computer science that gives the ability of machines to learn without being explicitly programmed. By using computational methods, machine learning has shown success in providing solutions for early stage diagnosis in a variety of medical domains. These methods are used to find hidden patterns from data and mine these data for decision- makers.

II. METHODOLOGY

This deals with the theoretical ability of the research work. It will give a piece of clear information about the concept of work. For the study, the dataset is collected form an online data repository. This data is cleaned using several preprocessing techniques. The feature selection is done on the dataset obtained from the repository. Once the data is

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105158 |

cleaned and processed, it is divided in training set and test set data. Four machine learning classification algorithms are trained using the training data. After the algorithms are train, they are implemented on the test data to obtain prediction.

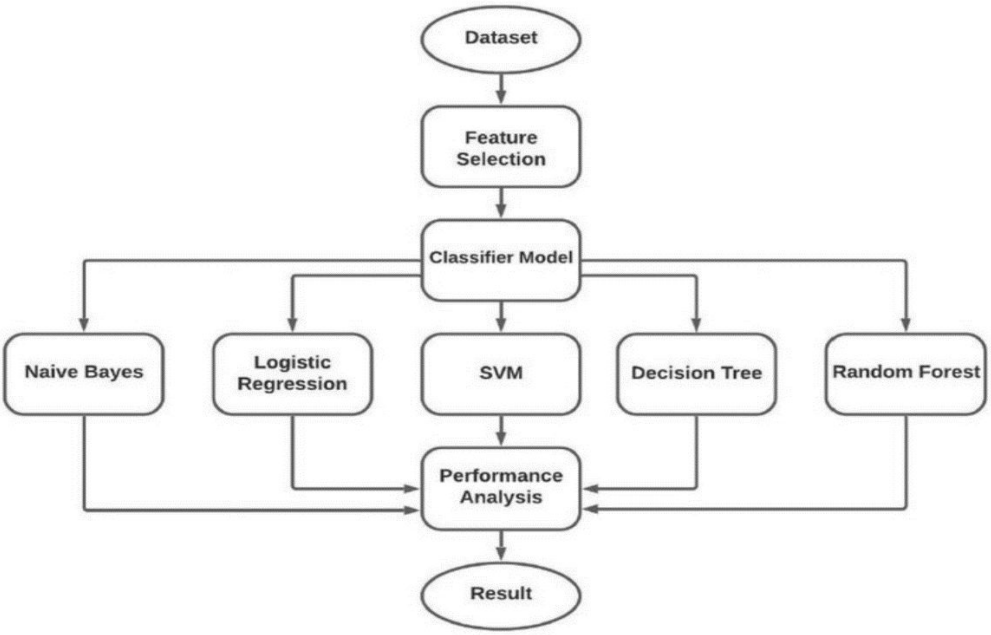

Figure 1: Methodology

III. MODELING AND ANALYSIS

Theprocessofstudyingaprocedureorbusinesstoidentifyitsgoalsandpurposesandcreatesystemsand procedures that will achieve them in an efficient way. Another view sees system analysis as aproblem-solving technique that breaks down a system into its component pieces for the purpose ofthestudying how well those componentparts work andinteractto accomplish theirpurpose.

XGBoostClassifier :

XGBoost is an optimized distributed gradient boosting library designed to be highly **efficient**, **flexible** and **portable**. It implements machine learning algorithms under the Gradient Boosting framework. XGBoost provides a parallel tree boosting (also known as GBDT, GBM) that solve many data science problems in a fast and accurate way. The same code runs on major distributed environment (Hadoop, SGE, MPI) and can solve problems beyond billions of examples. XGboost owns the ability to handle both types of situations, whether you need to go with regression or classification modelling.

Regression Algorithm: Suppose you have a certain amount of input features. Suppose your ML modelling target is to find such output variables that are continuous in nature and dependent on the considered output feature. In that case, the algorithms you use will be a regression one. In the case of regression, it's your responsibility to coach the data in such a way so that the ML model you will be designed based on the specified algorithm can evaluate the expected features of fresh datasets.

Classification Algorithm: Suppose you have multiple data sets, each holding data with lots of different features. But some of the features hold few similarities to their operations, output goals, etc. We can say there may seem several subsets of data in easier words that show some similar features. If you need to identify such kind of featured sub-sets of data to reach your expected output, then what you need to do is classify those data subsets based on feature similarities. But this type of classification will be automated. To initiate that, all of the data has to be coached based on the variables observations of the considered dataset. Once the data gets trained, it becomes active to classify or categorise all the upcoming new datasets based on its previous training.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105158 |

One of the key advantages of XGBoost is its ability to handle missing data and large datasets efficiently. It also has a number of hyperparameters that can be tuned to improve model performance, including the learning rate, depth of the trees, and regularization parameters.

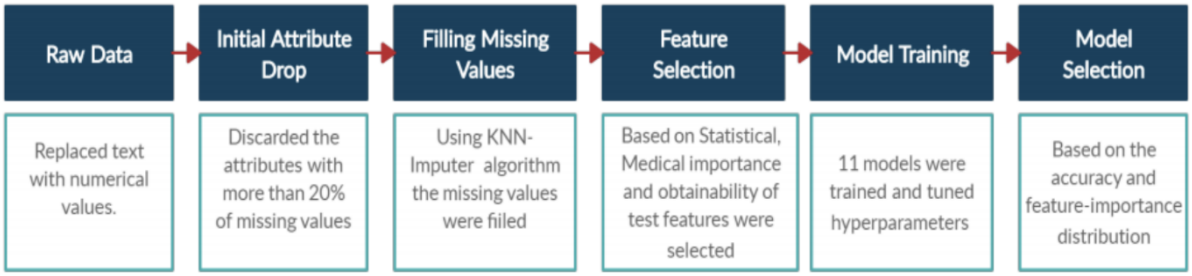

Figure 2: Proposed Layers

IV. RESULTS AND DISCUSSION

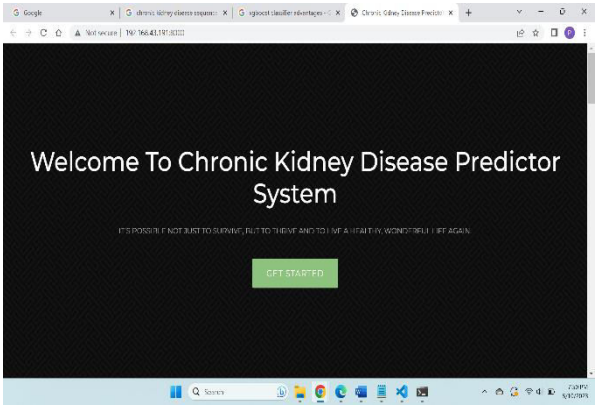

Figure 3: Home Page

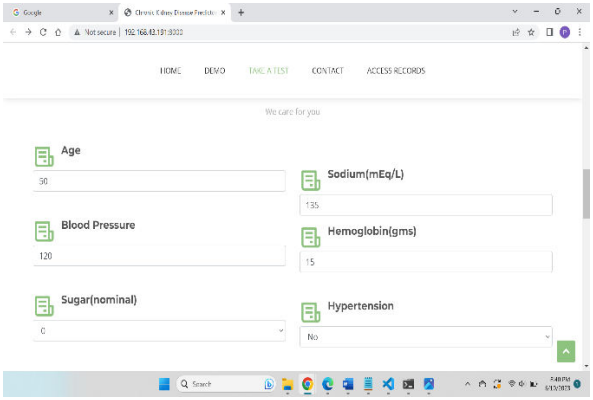

Figure 4: Test Page_1

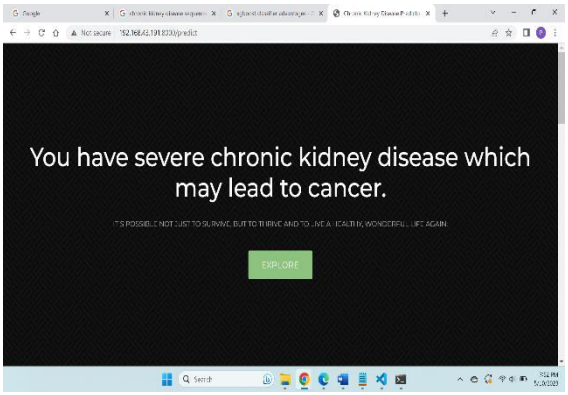

Figure 5: Result Page_1

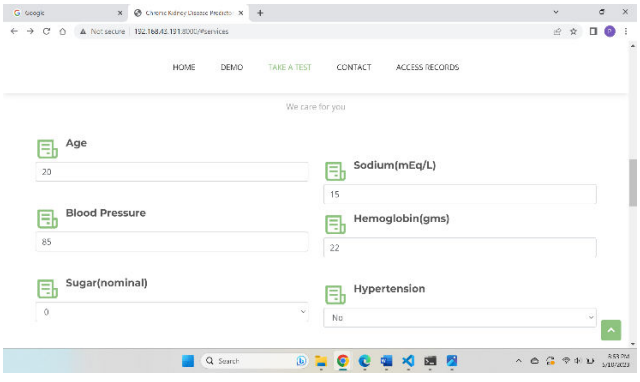

Figure 6: Test Page_2

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105158 |

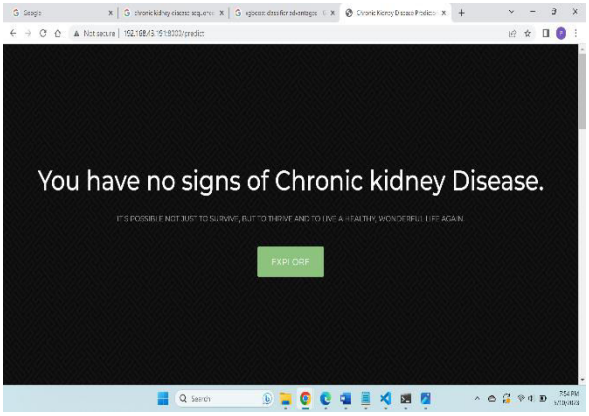

Figure 7: Result Page_2

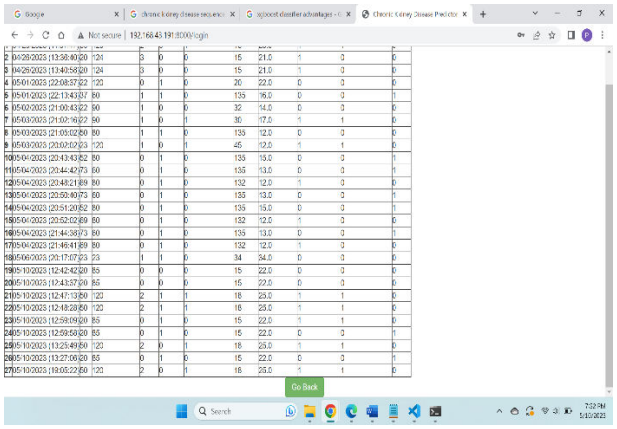

Figure 8: Database Record

Home Page: It isa introductory or startup page to the user.
Input Page_1: Here you have to input data for each attribute.
Result Page_1: It shows that you are positive to CKD based on the attributes.
Input Page_2: Here you have to input data for each attribute.
Result Page_2: It displays the result that you don't have CKD.
Database Record: It contains the information of each patients.

V. CONCLUSION

From the results, we can conclude that data preprocessing and Machine Learning combination is the right approach when we are trying to build a system to detect and classify Chronic Kidney Disease Present or Not.Weachieved10%increaseinaccuracycomparedtoourbasepaper.Thiscanbebecauseof the different approach we took while solving the problem. Through literature survey, we foundthatthemachinelearningapproachwasthebestwaytopredict whetherapersonisaffectedfromCKD or not. Majority of the papers used different algorithms such as Support Vector Machine (henceforth SVM), AdaBoost (henceforth AB), Linear Discriminant Analysis (henceforth LDA), and Gradient Boosting and so on. Most of the papers collected data set from hospitals and open-sourcerepositories. Toconclude machine learningapproach wouldyieldus the highestaccuracy.

REFERENCES

[1] Optimization of Prediction Method of Chronic Kidney Disease Using Machine Learning Algorithmby Pronab Ghosh, F. M. Javed Mehedi Shamrat, Shahana Shultana and Sami Afrin and was published in the year 2020.
[2] Classification of Chronic Kidney Disease using Logistic Regression, Feedforward Neural Network and Wide & Deep LearningBy Abdullah Al Imran, Md Nur Amin, Fatema TujJohora was publishedintheyear2018.
[3] A Machine Learning Methodology for Diagnosing Chronic Kidney DiseaseByJiongming Qin, and Lin Chen and Yuhua Liu was published intheyear2019.
[4] Chronic Kidney Disease Prediction Using Machine Learning MethodsbyImesh Udara Ekanayake andDamayanthi Herath was publishedin the year2020.
[5] Chronic Kidney Disease Prediction using Machine Learning Ensemble Algorithm by Nikhilapublished in the year2021.
[6] Preemptive Diagnosis of Chronic Kidney Disease Using Machine Learning Techniques by Reem A. Alassaf, Khawla A. Alsulaim , Noura Y. Alroomi, Nouf S. Alsharif, Mishael F. Aljubeir was published in the year 2018.
[7] Detection of Chronic Kidney Disease Using Machine Learning Algorithms with Least Number of Predictors by Marwa Almasoud, Tomas E Ward was published in the year 2019.
[8] Prediction of Chronic Kidney Disease – A Machine Learning Perspectiveby Pankaj Chittora, Sandeep Chaurasia, Gaurav Kumawat and Vadim Bolshev was Published in the year2021.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.379

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details