IJIRCCE

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

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 11, Issue 3, March 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 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104187 |

Chronic Kidney Disease Prediction Using Machine Learning

Vedant Bhoir, Saurav Sangle, Prof. Rasika Shintre

B.E. Student, Department of Computer, Smt. Indira Gandhi College of Engineering Navi Mumbai, Maharashtra, India

B.E. Student, Department of Computer, Smt. Indira Gandhi College of Engineering Navi Mumbai, Maharashtra, India

Project Guide, Department of Computer, Smt. Indira Gandhi College of Engineering Navi Mumbai, Maharashtra, India

ABSTRACT: Habitual order complaint(CKD) is an exceeding health problem that's affecting 10 of the world's population and 17 of the Indian population. So it's more important to diagnose it rightly. As we know that the machine literacy(ML) ways have unique advantages and that's why they're extensively used for analysis and vaticination of conditions. Experimenters have used the CKD dataset handed by UCI depository to develop prophetic models for CKD. In this website we've used LightGBM, KNN and decision tree algorithm. The LightGBM is chosen as a farther prophetic model for its high performance of speeding up the training time and lower memory consumption. The dataset has 24 attributes and requires 18 plus tests to be conducted to conclude these attributes which are expensive. That's why we've also reviewed some of the point selection styles and tried to use only the most important prophetic attributes to save the training time of model and the most important tried to reduce the cost needed to do these tests. Keywords K-Nearest Neighbours Algorithm KNN), GUI Graphical stoner Interface, Light grade Boosting Machine(LGBM).

KEYWORDS: **:** K-Nearest Neighbours Algorithm (KNN), GUI : Graphical User Interface, Light Gradient Boosting Machine (LGBM).

I. INTRODUCTION

Preface in this design, we present machine literacy ways for prognosticating habitual order complaint using clinical data. Machine learning system(algorithm) is used in the neural network. These prophetic models are constructed from habitual order complaint dataset and the performance of these models are compared together in order to elect the stylish classifier for prognosticating the habitual order complaint. First, we prognosticated if cases develop severe ckd, both including and banning information about the time it passed or date of the last visit. Our styles achieved top mean matthews correlation measure(mcc) of0.499 in the former case and a mean mcc of0.469 in the ultimate case also, we performed a point ranking analysis to understand which clinical factors are most important age, egfr, and creatinine when the temporal element is absent; hypertension, smoking, and diabetes when the time is present. We also compared our results with the current scientific literature, and bandied the different results attained when the time point is barred or included. Our results show that our computational intelligence approach can give perceptivity about opinion and relative important of different clinical variables that else would be insolvable to observe. In this website stoner need to fill biomedical data information asked by the system also website prognosticate whether the stoner has habitual order complaint ornon-chronic order complaint. We've used light gbm, knn and decision tree algorithm. The algorithm predicts whether the stoner has habitual order complaint ornon-chronic order complaint

II. LITERATURE SURVEY

We give an overview of some current exploration Among the rearmost technologies is computer automated order complaint discovery. Acute order Injury(AKI) is common and dangerous complaint in rehabilitated cases. It's associated with poor issues similar as a drop chance of survival, longer sanitarium stays and an increase progression of habitual order complaint. To opinion AKI, the KDIGO clinical practice guideline has been published for furnishing standardized criteria of Date 2023-04-29 Words 854 Characters 5803 Page 1 of 2 AKI description and the recommendation in medical pathway. also, early discovery of AKI in case at threat can also ameliorate the issues. This paper presents an approach to help the croaker in opinion and decision timber process. Simple wain and J48 were named as the algorithms for this process. Second, a conception of tool conditions and design named " AKI Helper " is

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

|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104187 |

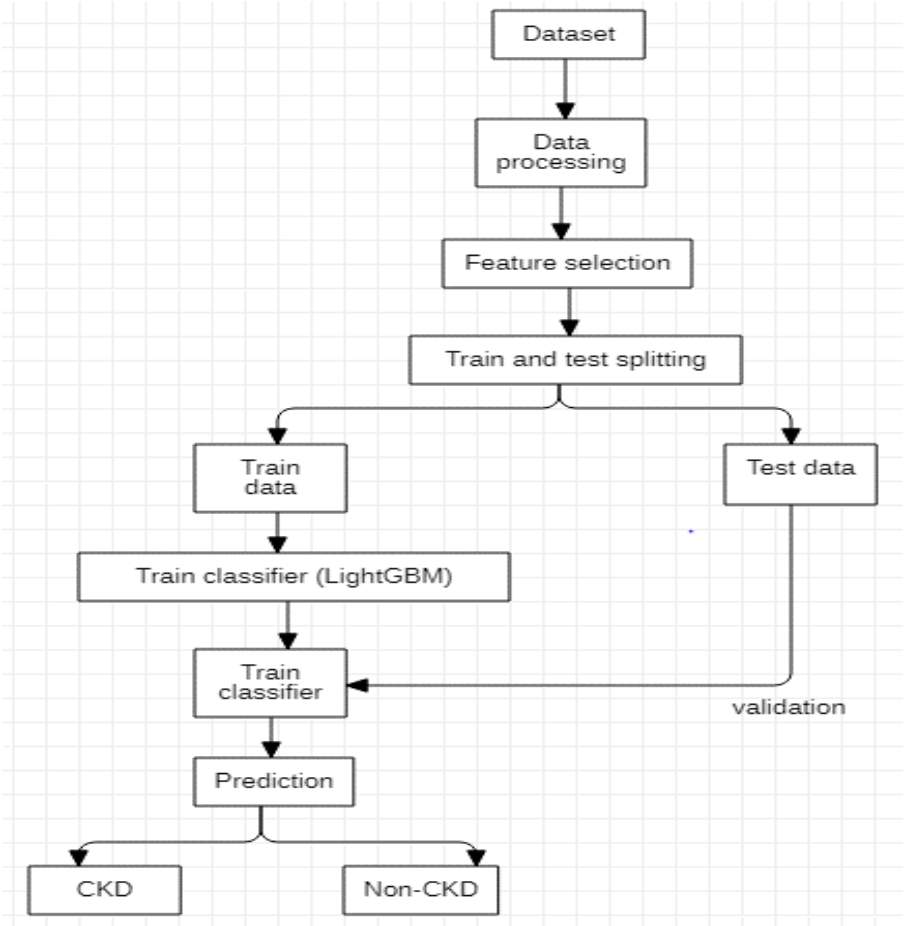
presented. This tool is created grounded on KDIGO guideline which is anticipated to use for opinion and staging inflexibility of AKI. The clinical practice guidelines for defining AKI've been published in colorful delineations similar as RIFLE, AKIN(Acute order Injury Network) and KDIGO(order complaint perfecting Global issues). still, the prevalence rate of AKI is still adding because it isn't a single complaint. The progression of AKI comes from multiple clinical conditions that make the individual of this complaint come more complex. likewise, the perpetration of these guidelines isn't successful as it should be. This is because of the difference in shoes from multidisciplinary stakeholders, walls in knowledge and a lack of understanding how to apply these guidelines in clinical practice. point rankings styles grounded on Random timbers are among the most effective ways(60),(61), particularly in the environment of bioinformatics(62),(63) and health informatics(64). Since Random timbers attained the top vaticination scores for double bracket, we concentrate on this system for point ranking..

III. PROPOSED ALGORITM

We proposed a machine learning based diagnosis method for identification of chronic kidney disease in this website. We reviewed XGBoost algorithm and compared with LightGBM which is lighter version of the Gradient Boosting Decision Tree. The LightGBM is chosen as a further predictive model for its high performance of speeding up the training time and less memory consumption.

- Firstly we reviewed lightGBM model, then used for chronic kidney classification problem. LightGBM is an open-source framework for gradient boosted machines.
- It requires very less training time as compared to all machine learning algorithms and its also faster than the other gradient boosting algorithms such as Xgboost,Catboost etc.
- Secondly, we are applying some of the feature selection methods from filter, wrapper, and embedded approach by combining all results we are selecting 5 attributes which are common in the results of all methods.

Lastly, we are applying CKD-EPI equation to find out the stage of ckdpatient.Serum creatine which is considered as a important indicator of CKD is used to compute the stages of CKD by putting it into the CKD-EPI equation



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 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104187 |

Fig. 1 System Architecture

We're using The nimble methodology promotes the nonstop commerce of the development and testing during the SDLC process of any design. Unlike the cascade model, the development and testing conditioning in the nimble model are contemporaneous. nimble methodology allows important communication between the guests, inventors, testers, and directors. In the nimble system, the entire design is divided into small incremental shapes. All of these shapes are handed in duplications, and each replication lasts from one to three weeks. nimble development methodology and testing practices have worked prodigies for several associations with positive aspects. The positive aspects of nimble aren't hidden. They're veritably much visible in associations.

1) Agile Model

Agile has been used in learning and teaching for a while now, thanks to earlier research that showed how agile principles in business and education generally relate to one another. Consideration should be given to the ongoing and ingrained practise of feedback and reflection as a foundation for the development of some essential professional talents, capabilities, and expertise for desired graduate employability traits. Agile approaches may improve teamwork. PBL and positive emotions are connected to the emphasis that agile principles place on their results. Time-boxed phases promote a learning pace that is sustainable at the technique level with frequent feedback and evaluation on actual learning results. By promoting students' self-management, this approach can help increase confidence and decrease uncertainty.

2) (KNN)

The k-nearest neighbour's 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 an individual data point.

3) Tensor Flow

The preferred TensorFlow is an open-source software library for numerical computation using data flow graphs. The graph nodes represent mathematical operations, while the graph edges represent the multidimensional data arrays (tensors) that flow between them. This flexible architecture enables you to deploy computation to one or more CPUs or GPUs in a desktop, server, or mobile device without rewriting code. TensorFlow provides stable Python and C APIs as well as nonguaranteed backward compatible API for C++, Go, Java, JavaScript, and Swift.

4) WorkingTensor Flow

TensorFlow allows developers to create dataflow graphs— structures that describe how data moves through a graph, or a series of processing nodes. Each node in the graph represents a mathematical operation, and each connection or edge between nodes is a multidimensional data array, or tensor. Tensor flow is a powerful machine learning library to create models and neural networks. So, before we start What are Artificial neural networks? Here is a simple and clear definition of artificial neural networks.

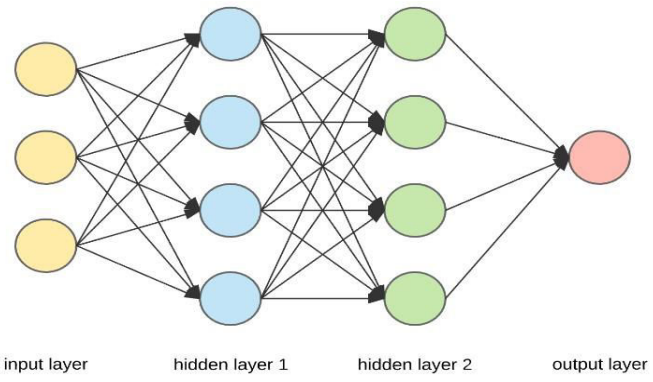


Fig. 3 Tensor flow model

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 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104187 |

5) Light Gradient Boosting Machine (LGBM)

Light Gradient Boosting Machine referred as Light GBM is a gradient boosting framework. It is a decision tree based learning algorithm, mostly used for machine learning classification problems. It can also be applied on machine learning based regression tasks. It takes very less computational time for training of data as compared to other gradient boosting algorithms as the trees in LightGBM algorithm grows leaf-wise while in other algorithms it grows level-wise.

Light GBM uses two important sampling methods such as GOSS and EFB,

- With GOSS the model excludes the instances on the basis of gradients, the instances with large gradients which contributes more to the information gain are kept and it excludes instances with small gradients based on some predefined threshold.
- EFB reduces the features to reduce the training complexity which are basically mutually exclusive. Then these exclusive features later on bundled into a single feature called Exclusive feature bundle and treated as a single feature.

A) GOSS Calculation

1) Keep all the large gradient instances.
2) Randomly sample small gradient instances
3) Use a constant multiplier for small gradient data instances during information gain computation in the tree building process.
4) If a be the instances with large gradients and b is the randomly sampled small gradient instances, the sampled data will be amplified by (1-a/b).

B) EFB

1) Find out the mutually exclusive features which can be bundled together.
2) Merge the features into a single feature, resulted in the first step.

C) Decision tree algorithm

A decision tree is a non-parametric supervised learning algorithm, which is utilized for both classification and regression tasks.

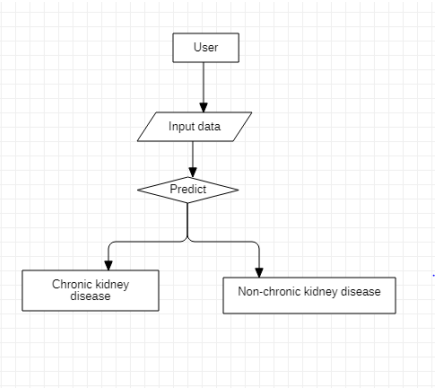


Fig. 4 Data Flow Chart

IV. IMPLEMENTATION

INPUT

Launch the Anaconda software, navigate to the Node.js file in the command prompt, and launch the program file for disease identification. The program began to function, and using the parameters input, it recognized the result of disease and produced statistics for kidney disease if any are there. There are the parameters which patients or doctor has to given to program

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

|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104187 |

Attribute Number	CKD Attributes	Data type
1	Age	Number (in years)
2	Blood Pressure	Number (in mm/hg)
3	Specific Gravity	Nominal
4	Albumin	Nominal(1-5)
5	Sugar	Nominal(1-5)
6	Red Blood Cells	Nominal (normal,abnormal)
7	Pus Cells	Nominal (normal,abnormal)
8	Pus Cell Clumps	Nominal (normal,abnormal)
9	Bacteria	Nominal (normal,abnormal)
10	Blood Glucose Random	Number (in mgs/dl)
11	Blood Urea	Number (in mgs/dl)
12	Serum Creatinine	Number (in mgs/dl)
13	Sodium	Number (in mgs/dl)
14	Pottasium	Number (in mgs/dl)
15	Hemoglobin	Number (in gms)
16	Packed Cell Volume	Nominal
17	White Blood Cell Count	Number (in cells/cumm)
18	Red Blood Cell Count	Number (in cells/cumm)
19	Hypertension	Nominal (yes, no)
20	Diabetes Mellitus	Nominal (yes, no)
21	Coronary Artery Disease	Nominal (yes, no)
22	Appetite	Nominal (good, poor)
23	Pedal Edema	Nominal (good, poor)
24	Anemia	Nominal (yes, no)

V. SIMULATION RESULTS

This parameter which are given to the program it will predict on the basis of parameters which are compared to database for similar or close to result disease which will be scale from 0 to 10 to show how severe or which type of disease is there to operate or treat as soon possible.

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 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104187 |

This proposal is mainly used for the recorded patient's health through 24 aspects of body which were in diagnosis that is our input. It is a way of generating a report on the range of 0 to 10. With this feature, the diseases can be identified . This helps out the doctor to do direct procedure or treatment immediately , this will increase the response time of doctor to act on or check on patient's kidney for right procedure .this will also help patient to get right information through there diagnosis stats that how severe it can be or it is.

IV. CONCLUSION

From this paper we come to know that - More than 5% of adults suffer from different types of kidney disease, and millions of people die prematurely from cardiovascular diseases associated with chronic kidney disease (CKD) in each year.

CKD is a gradual loss in kidney function causes kidney failure, if it is detected as early as possible then by taking preventive measures against it we can completely cure it or we can prevent the progression of CKD in patients. In this study, we have developed a machine learning based diagnosis system for diagnosis of chronic kidney disease.We have used LightGBM model for this classification. This model gave us 100% accuracy in model training and also took very less training time as compared to other machine learning algorithms.Also we did features selection, we have selected important features to reduce the training time and as the all attributes requires 18 plus test to be conducted which are costly, by selecting important attributes the cost of the test is also reduced.

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

|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104187 |

V. FUTURE SCOPE

Our project can be improvised by integrating it with feature selection methods to select the best subset features to develop the models. It is better to see the difference in performance results using unsupervised or deep learning algorithms models. The proposed model supports the experts to give the fast decision, it is better to make it a mobile-based system that enables the experts to follow the status of the patients and help the patients to use the system to know their status.

REFERENCES

[1] Adeola Ogunleye and Qing-Guo Wang[1] developed Extreme Gradient Boosting (XGBoost) based Chronic Kidney disease diagnosis system.They reviewed existing classification algorithmsand have used (XGBoost) model by combining three features selection techniques for rapid and accurate diagnosis of CKD.

[2] C. S. Lee and M. H. Wang [2]designed a fuzzy expert system for diabetes prediction. They developed a unique five layer fuzzy ontology to vesion the area understanding with uncertainty and volume the bushy ontology to the diabetes area. Additionally they worked on a semantic decision support agent[SDSA] for semantic decision making in diagnosis of diabetes.

[3] Deepa Gupta , Sangita Khare and Ashish Aggarwal [3] proposed a method for the prediction of diagnostic codes for chronic diseases by applying machine learning algorithms. They made a specialty of scientific records, claims records for analysing 11 continual disorder which includes kidney disorder, osteoporosis, arthritis etc. the use of the declared records.

[4] V.Krishnaiah, Dr.G.Narsimha, Dr.N.Subhash Chandra [4] proposed a lung cancer diagnosis system with the help of data mining type techniques.They compared different machine learning classification techniques such as Naïve Bayes, Decision Tree and Neural Network. In that Naïve Bayes performed better than others as it picks out allo of the considerable predictors.

[5] Younghwan Shin, Hyun Soo Chung, SangdoKim ,Sang Gil Han, Jong Moon Chung, Junho Cho [5] proposed a unique Blood sample-based Emergency department (ED) Return (BER) systemwhich is usedfor the prediction of the ED return. Here they have used LightGBM algorithm against XGBoost algorithm. They in this comparison got effective outcomes as compared to xgboost. LightGBM effectively predicted the ED return probability of hospitals.

[6] GuolinKe , Qi Meng , Thomas Finley , Taifeng Wang , Wei Chen , Weidong Ma , QiweiYe Tie-Yan Liu[6], developed a fast, an accurate and highly efficient algorithm called LightGBM.They have included GOSS and EFB sampling methods in LightGBM for boosting up the performance of LightGBM. LightGBM appreciably and effectively outperformed than XGBoost and SGB in the contrast of training speed and consumption of memory.

[7] Fatimah Alzamzam, Mohmad Hoda and Abdul motalel EL soddik[7], usedLight gradient boosting machine (LightGBM) for sentiment analysis and classification based on short texts. Their LGBM model was trained to classify tweets sentiments in: positive,negative or neutral categories. Result showed that, LightGBM based LGBM sentiment classifier outperformedthan the other classification algorithms in case of accuracy and F-scores.

[8] DinglingGe ,shunyuchang [8] proposedaLightGBM based Credit Card Fraud Detection system.They compared various algorithms such as SVM (Support Vector Machine), Random Forest and XGBoost. The results shown that LightGBM is an effective modelling algorithm than the above.

[9] Jian Ping Li, Amin Ul Haq, Salah Ud Din, Jalaluddin Khan, Asif Khan [9], presented heart disease diagnosis using machine learning techniques. They reviewed various classification algorithms includes Support vector machine, Logistic regression, Artificial neural network, K-nearest neighbor, Naïve bays, and Decision tree.Also they took a look on some of the standardized feature selection methods such as Minimal redundancy maximal relevance, Least absolute shrinkage selection operator and Local learning to remove irrelevant features and save the training time of model. They also proposed a unique and fast conditional mutual information feature selection algorithm.

[10] Jović , K. Brkić and N. Bogunović [10],presented feature selection and has provided an overview of the available feature selection techniques to handle several different typesof problems. Alsoanalyzed application of feature selection domain and reviewed comparative studies on feature selection to analyzed which method is best suited for the particular class of problem

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

|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104187 |

BIOGRAPHY

Vedant Bhoir
is pursuing the Bachelor degree (B.E.) in Computer Engineering from Smt. Indira Gandhi college Of Engineering ,Navi Mumbai.

Saurav Sangle
is pursuing the Bachelor degree (B.E.) in Computer Engineering from Smt. Indira Gandhi college Of Engineering ,Navi Mumbai

PROF. Rasika Shintre
Obtained the Bachelor degree (B.E. Computer) in the year 2011 from RamraoAdik Institute of Technology (RAIT), Nerul and Master Degree (M.E. Computer)From Bharti Vidyapeeth College Of Engineering, Navi Mumbai. She is Asst. Prof in Smt. Indira Gandhi college Of Engg. Of Mumbai University and having about 11 years of experience. Her area of interest include Data Mining and Information Retrieval.

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