IJIRCCE

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

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 3, March 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.488

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 3, March 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0903160 |

Prediction of Chronic Kidney Disease Using KNN and Naive Bayes Classifier

Mohammad Jaffar A , G.Yuvaraj M.E.,

UG Student, Department of Computer Science and Engineering, Velammal Institute of Technology, Chennai, Tamil Nadu, India

Assistant Professor, Department of Computer Science and Engineering, Velammal Institute of Technology, Chennai, Tamil Nadu, India

ABSTRACT: Chronic kidney disease (CKD) is a worldwide medical condition with high grimness and mortality rate, and it instigates different illnesses. Early discovery of CKD empowers patients to get opportune treatment to improve the movement of this infection. Machine learning models can successfully help clinicians accomplish this objective because of their quick and exact acknowledgment execution. KNN attribution was utilized to fill in the missing qualities, which chooses a few complete examples with the most comparable estimations to measure the missing information for each inadequate example. Missing qualities are generally found, in actuality, clinical circumstances since patients may miss a few estimations for different reasons. After viably rounding out the inadequate informational index, Machine learning calculations like k-nearest neighborand Naive Bayes classifier were used to establish models. we estimated that this procedure could be material to more muddled clinical information for infection determination.

KEYWORDS: CKD, Naïve bayes , KNN

I. INTRODUCTION

The aim of the proposed work is to identify the chronic kidney disease as early as possible by using machine learning methods. Five iterations of 10-fold cross-validation were conducted for each classifier using the CKD dataset, showing stability. The models showed to be stable because the surrogate models yield the same prediction for the same input data. The performance metrics used to compare the classifications of the most-experienced nephrologist and the machine learning algorithms based on the hypertension, DM, creating, urea, albuminuria, age, gender, and GFR attributes. More specifically, the CCI, ICI, MAE, RMSE, TP rate, FP rate, P, ROC area, and PRC performance metrics were applied. Presents a complete comparison, considering the results of the classifiers and the opinions of the three nephrologists. As expected, the k values clearly increase when the most-experienced nephrologist is included in the comparisons. The KNN approach is used to impute missing values. The assumption employed in imputation operation is that similar instances would have similar characteristics. After evaluating the above models, the potential component models were extracted for misjudgment analysis to determine which would be used as the components. The misjudgment analysis here refers to find out and compare the samples misjudged by different models, and then determine which model is suitable to establish the final integrated model. The proposed CKD diagnostic methodology is feasible in terms of data imputation and samples diagnosis. After unsupervised imputation of missing values in the data set by using KNN imputation, the integrated model could achieve a satisfactory accuracy.

II. LITERATURE SURVEY

In the year 2020 Adeola Ogunleye and Qing-Guo Wang proposed a method. [1] the XGBoost method has been studied and optimised for CKD diagnosis. The resulting CKD models are compared with the existing CKD models in the domain. The proposed full model has achieved an accuracy, sensitivity and specificity of 1.000, 1.000 and 1.000, respectively. Three feature selecting techniques are combined by leveraging the strengths of each technique. A reduced model with about a half of the full features has an accuracy, sensitivity and specificity of 1.000, 1.000 and 1.000, respectively.

[2]In the year 2018, M. Mahyoub, M. Randles, T. Baker, and P. Yang proposed a method ,The feasibility of two in-house fuzzy classifiers, fuzzy rule-building expert system (FuRES) and fuzzy optimal associative memory (FOAM), for diagnosis

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 3, March 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0903160 |

of patients with chronic kidney disease (CKD) was investigated. A linear classifier, partial least squares discriminant analysis (PLS-DA), was used for comparison. The CKD data used in this work were taken from the UCI Machine Learning Repository. . In [3], the year 2019 Gabriel R. Vásquez-Morales, Sergio M. Martínez-Monterrubio proposed a method Explainable Prediction of Chronic Renal Disease in the Colombian Population using Neural Networks and Case-Based Reasoning where Optimization algorithm,despite the demonstrated efficiency of the neural networks to predict CKD, this machine-learning paradigm is opaque to the expert regarding the explanation of the outcome. It would conduct a user based evaluation to validate the impact of the generated explanations in the acceptance by the expert of the predictions given by the neural network

In 2019 [4] Navaneethan Bhaskar and Suchetha M proposed a method Feature Extraction and Classification a new sensing technique for the automated detection of kidney disease. The salivary urea concentration is monitored to detect the disease. A new sensing approach is introduced to monitor the urea levels in the saliva sample.We have performed the statistical analysis to determine how well the proposed sensing method values and traditional urea estimation values are correlated.

III. PROPOSED METHODOLOGY AND DISCUSSION

CHRONIC kidney infection (CKD) is a worldwide public medical condition influencing roughly 10% of the total populace. The level of commonness of CKD in China is 10.8%, and the scope of commonness is 10%-15% in the United States. As indicated by another study, this rate has arrived at 14.7% in the Mexican grown-up all inclusive community.CKD doesn't show evident side effects in its beginning phases. Along these lines, the sickness may not be identified until the kidney loses about 25% of its capacity.To overcome the past CKD analytic models, we locate that the vast majority of them experiencing either the strategy utilized to attribute missing qualities has a restricted application range or generally low exactness. Accordingly, in this work, we propose a technique to broaden application scope of the CKD analytic models. Simultaneously, the precision of the model is additionally improved.

The proposed CKD symptomatic procedure is achievable in terms of information attribution and tests finding. After solo attribution of missing qualities in the informational index by utilizing KNN ascription, the coordinated model could accomplish an acceptable exactness. Consequently, we theorize that applying this technique to the functional finding of CKD would accomplish an alluring impact. What's more, this procedure may be material to the clinical information of different infections in real clinical determination. In any case, during the time spent building up the model, because of the constraintsof the conditions, the accessible information tests are generally little, including as it were 400 examples. In this way, the speculation execution of the model may be restricted. Also, due to there are as it were two classifications (ckd and notckd) of information tests in the information set, the model cannot analyze the seriousness of CKD. In the future, countless more unpredictable and delegate information will be gathered to prepare the model to improve the speculation execution while empowering it to distinguish the seriousness of the illness. We accept that this model will be increasingly more wonderful by the expansion of size and nature of the information. We propose a methodology to extend application range of the CKD diagnostic models. At the same time, the accuracy of the model is further improved.

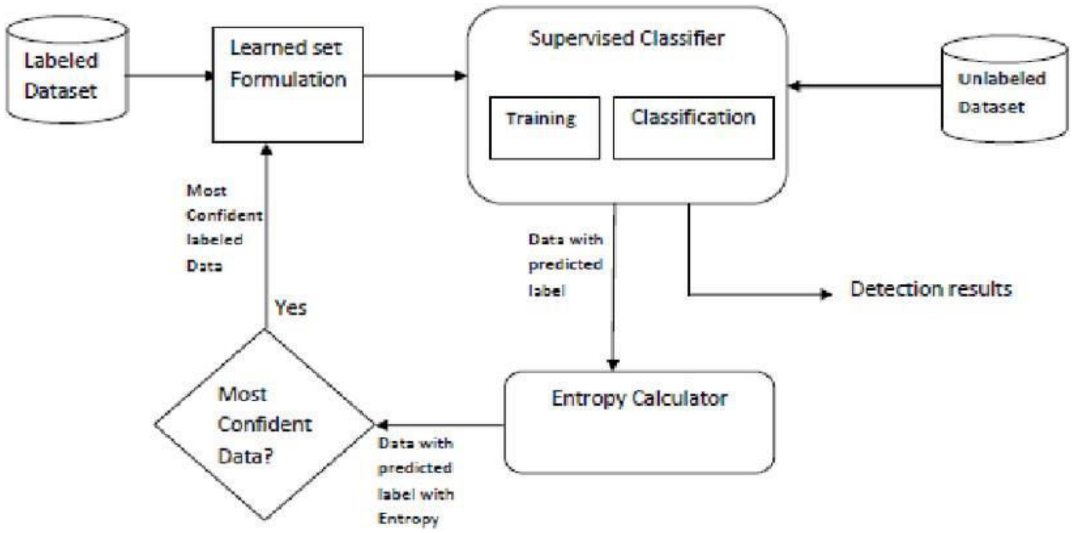

Figure 1- Architecture Diagram

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 3, March 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0903160 |

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.The dataset used in thisChronic kidney disease dataset taken from UCI:https://archive.ics.uci.edu/ml/datasets/chronic_kidney_disease

Data Preparation:

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

Model Selection:

It is a supervised learning algorithm that includes more dependent variables. The response of this algorithm is in the binary form. Logistics regression can provide the continuous outcome of a specicfi data. This algorithm consists of statistical modelwith binary variables .

Saving the Trained Model:

Take trained and tested model into the first step is to save it into a .h5 or .pkl file using a library like pickle , installed.import the module and dump the model into . pkl file

IV. EXPERIMENTAL RESULTS

Figure 2- login Page

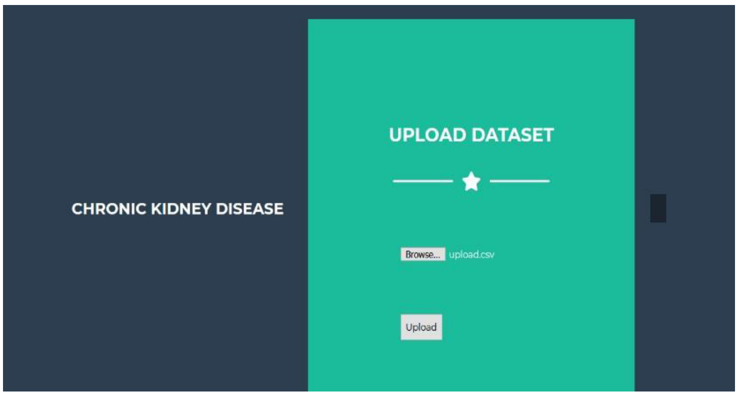

Figure 3- Dataset upload page

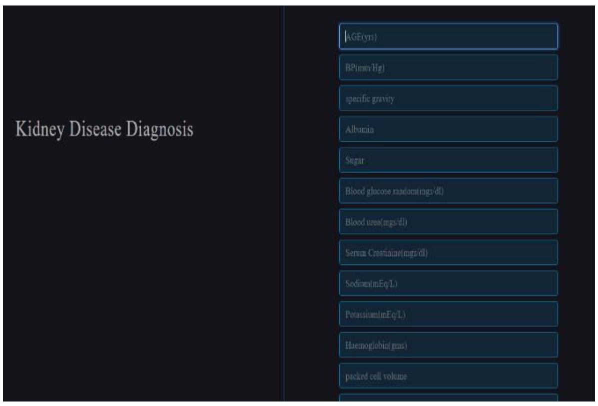

Figure 4.CKD prediction page

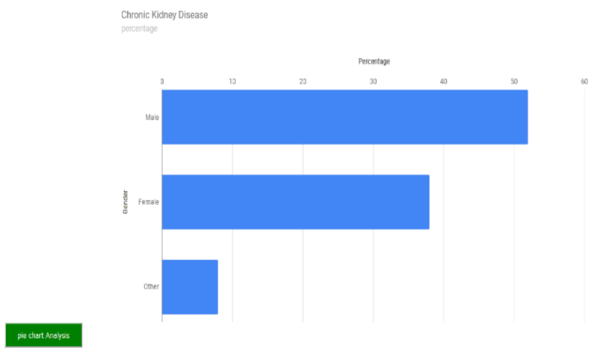

Figure 5. Result visualization

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 3, March 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0903160 |

V. CONCLUSION AND FUTURE ENHANCEMENT

The proposed CKD diagnostic methodology is feasible in terms of data imputation and samples diagnosis. After unsupervised imputation of missing values in the data set by using KNN imputation, the integrated model could achieve a satisfactory accuracy. Hence, we speculate that applying this methodology to the practical diagnosis of CKD would achieve a desirable effect. In addition, this methodology might be applicable to the clinical data of the other diseases in actual medical diagnosis. However, in the process of establishing the model, due to the limitations of the conditions, the available data samples are relatively small, including only 400 samples. Therefore, the generalization performance of the model might be limited. In addition, due to there are only two categories (ckd and notckd) of data samples in the data set,the model cannot diagnose the severity of CKD.

In the future, an enormous number of more perplexing and delegate information will be gathered to prepare the model to improve the speculation execution while empowering it to recognize the seriousness of the sickness. We accept that this model will be increasingly more amazing by the expansion of size and nature of the information.

REFERENCES

1. Adeola Ogunleye and Qing-Guo Wang,`` XGBoost Model for Chronic Kidney Disease Diagnosis", IEEE/ACM transactions on computational biology and bioinformatics, vol. 17, no. 6, november/december 2020
2. Subasi, E. Alickovic, and J. Kevric, ``Diagnosis of chronic kidney disease by using random forest," in Proc. Int. Conf.Med. Biol. Eng., Mar. 2017, pp. 589594.
3. L. Zhang, ``Prevalence of chronic kidney disease in China: A crosssectional survey," Lancet, vol. 379, pp. 815822, Mar.2012.
4. Singh, G. Nadkarni, O. Gottesman, S. B. Ellis, E. P. Bottinger, and J. V. Guttag, ``Incorporating temporal EHR data in predictive models for risk stratication of renal function deterioration," J. Biomed. Informat., vol. 53, pp. 220228, Feb. 2015.
5. M. Cueto-Manzano, L. Cortés-Sanabria, H. R. Martínez-Ramírez, E. Rojas-Campos, B. Gómez-Navarro, and M. Castillero-Manzano, ``Prevalence of chronic kidney disease in an adult population," Arch. Med. Res., vol. 45, no. 6, pp. 507513, Aug. 2014.
6. H. Polat, H. D. Mehr, and A. Cetin, ``Diagnosis of chronic kidney disease based on support vector machine by feature selection methods," J. Med. Syst., vol. 41, no. 4, p. 55, Apr. 2017.
7. Barbieri, F. Mari, A. Stopper, E. Gatti, P. Escandell-Montero, J. M. Martínez-Martínez, and J. D. Martín-Guerrero, ``A new machine learning approach for predicting the response to anemia treatment in a large cohort of end stage renal disease patients undergoing dialysis," Comput. Biol. Med., vol. 61, pp. 5661, Jun. 2015.
8. V. Papademetriou, E. S. Nylen, M. Doumas, J. Probsteld, J. F. Mann, R. E. Gilbert, and H. C. Gerstein, ``Chronic kidney disease, basal insulin glargine, and health outcomes in people with dysglycemia: The ORIGIN Study," Amer. J. Med., vol. 130, no. 12, pp. 1465.e271465.e39, Dec. 2017.
9. N. R. Hill, ``Global prevalence of chronic kidney diseaseA systematic review and meta-analysis," PLoS ONE, vol. 11, no. 7, Jul. 2016, Art. no. e0158765.
10. M. M. Hossain, R. K. Detwiler, E. H. Chang, M. C. Caughey, M.W. Fisher, T. C. Nichols, E. P. Merricks, R. A. Raymer,
11. M. Whitford, D. A. Bellinger, L. E. Wimsey, and C. M. Gallippi, ``Mechanical anisotropy assessment in kidney cortex using ARFI peak displacement: Preclinical validation and pilot in vivo clinical results in kidney allografts," IEEE Trans. Ultrason.,Ferroelectr., Freq. Control, vol. 66, no. 3, pp. 551562, Mar. 2019.
12. M. Alloghani, D. Al-Jumeily, T. Baker, A. Hussain, J. Mustana, and A. J. Aljaaf, ``Applications of machine learning techniques for software engineering learning and early prediction of students' performance," in Proc. Int. Conf. Soft Comput. Data Sci., Dec. 2018, pp. 246258.