IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

USE OF MACHINE LEARNING ALGORITHMS FOR EFFECTIVE CHRONIC KIDNEY DISEASE PREDICTION

B. Siva Jyothi [*1], G.V.Gayathri [*2]

[*1,*2]Assistant Professor, Department of Computer Science & Engineering(AIML-DS),

Anil Neerukonda Institute of Technology & Sciences,

Visakhapatnam, Andhra Pradesh, India.

ABSTRACT

Chronic Kidney Disease (CKD) is a global health issue with a high morbidity and mortality rate, as well as a risk factor for other diseases. Because there are no clear symptoms in the early stages of CKD, patients frequently miss the condition. Early identification of CKD allows patients to obtain timely treatment to slow the disease's progression. Because of their fast and precise recognition performance, machine learning models can successfully assist doctors in achieving this goal. In this proposed work, we suggest a machine learning methodology for diagnosing CKD in the early stages and to test the performance of proposed ML algorithms, we collected the CKD data set from the KAGGLE website, which contains a substantial number of missing values. Normally the dataset may contain some missing values and in order to fill those missing values we try to assume mean value with written type string. If these missing values are not identified then the outcome may be inappropriate. By conducting various experiments on our proposed dataset by taking the four machine learning algorithms (Logistic Regression, SVM, Random Forest Classifier, Decision Tree Classifier),we finally came to an conclusion that random forest outperforms the other machine learning models in terms of accuracy.

KEY WORDS:

Chronic Kidney Disease, Logistic Regression, SVM, Random Forest Classifier, Decision Tree Classifier.

1. INTRODUCTION

Chronic kidney disease (CKD) was ranked the 18th largest cause of death worldwide in 2010, according to the Global Burden of Disease Study (GBDS), up from 27th in 1990 [2]. Chronic kidney disease affects nearly 500 million individuals globally [3], [4] with a disproportionately high burden in developing countries, particularly South Asia and sub-Saharan Africa [5]. According to a 2015 study, there were 110 million persons with CKD in high-income nations (men 48.3 million, women 61.7 million), but 387.5 million in low- and middle-income countries [6]. The entire population with CKD is estimated to be 14 percent in a global assessment of six

areas, including Bangladesh [7]. Another study found a 26% incidence of chronic kidney disease among urban Dhaka inhabitants over 30 years old [8], while another found a 13% prevalence of chronic kidney disease among urban Dhaka residents over 15 years old [9]. A community-based prevalence research in Bangladesh in 2013 found that one-third of rural individuals were at risk of developing CKD, which was commonly misdiagnosed at the time [10]. The observed differences in CKD prevalence between Bangladeshi groups, on the other hand, could be explained by a variety of factors, including the cross-sectional research design with a small sample size, the study time, and the geographic location of the study.

The prevalence of CKD varies by age group, gender, socioeconomic level, and geographic region, according to one study [7]. Patients with chronic kidney disease (CKD) are more likely to develop end-stage renal disease (ESRD), which necessitates costly treatment options such as dialysis and kidney transplantation [11], and this financial burden leads to long-term medical and psychological troubles [12, 13]. Early detection of kidney dysfunction may aid with rectification, which is not always attainable. The term "responsibility" refers to the act of determining whether or not a person is responsible for his or her own actions. The primary goal of this work is to predict renal disease by analysing data from those indices and using three machine learning classification algorithms to forecast the disease, then selecting the approach with the best prediction accuracy. Nearest neighbours classifier, decision tree classifier (DT), and logistic regression are the three classification algorithms employed. Machine learning classifiers are used to forecast the class, target, labels, and categories of a data point. Classification is a sort of supervised learning in which input data is given to the targets. Among the applications are medical diagnosis, spam detection, and targeted marketing. They achieve this by employing a mapping function () to convert discrete input variables () into discrete output variables.

2. LITERATURE SURVEY

In this section we will mainly discuss about the background work that is carried out in order to prove the performance of our proposed Method. Literature survey is the most important step in software development process. For any software or application development, this step plays a very crucial role by determining the several factors like time, money, effort, lines of code and company strength. Once all these several factors are satisfied, then we need to determine which operating system and language used for developing the application. Once the programmers start building the application, they will first observe what are the pre-defined inventions that are done on same concept and then they will try to design the task in some innovated manner.

SI NO.	TITLE OF THE PAPER	AUTHOR NAME AND YEAR OF PUBLICATION	TECHNIQUE USED	MERITS	DEMERITS
1	Chronic Kidney Disease Prediction Using Machine Learning Techniques	Saurabh Pal ,2022	Here we applied three machine learning classifiers Logistic Regression (LR), Decision Tree (DT), and Support Vector Machine (SVM) for analysis and then used bagging ensemble method to improve the results of the developed model	Kidney Disease Collection is summarized by category and non-linear features. We get the best result in the case of decision tree with accuracy of 95.92%.	We want to extend the same algorithms in ensemble model to increase the accuracy.
2.	***Machine Learning Techniques for Chronic Kidney Disease Risk Prediction***	Elias Dritsas ,2022	In this paper, we apply class balancing in order to tackle the non-uniform distribution of the instances in the two classes, then features ranking and analysis are performed, and finally, several ML models are trained and evaluated based on various performance metrics.	The derived results highlighted the Rotation Forest (RotF), which prevailed in relation to compared models with an Area Under the Curve (AUC) of 100%, Precision, Recall, F-Measure and Accuracy equal to 99.2%.	It is not giving more accuracy for large dataset.
3.	CKD prediction using ML	Dibaba Adeba Debal, 2021	Early and accurate detection of the stages of CKD is believed to be vital to minimize impacts of patient's health complications such as hypertension, anemia (low blood	In this study, both binary and multi classification for stage prediction have been carried out.	This Multi Class takes more time to process and cross validation failed with

			count), mineral bone disorder, poor nutritional health, acid base abnormalities, and neurological complications with timely intervention through appropriate medications.		some ML algorithms.

3. EXISTING METHODOLOGY

In the existing system there was no proper method to identify the chronic kidney disease prediction using data mining algorithms. The following are the main limitations in the existing system.

LIMITATION OF EXISTING SYSTEM

1. More Time Delay in finding the route cause of kidney diseases.
2. There is no prevention technique due to late prediction.
3. There is no early prediction of chronic kidney disease.
4. There is no method to identify the kidney diseases using ML algorithms.

4. PROPOSED MODEL

In proposed system we are applying different Machine Learning Algorithms to detect Chronic Kidney disease. In this Project we used the features like age, sugar(su),blood pressure(bp),hyper tension(htn),pus cell(pu),RBC,WBC,Coronary Artery Disease(cad),etc…to classify CKD.After data preprocessing a very well cleaned data is input to these algorithms(Logistic regression, Support Vector Machine, Random forest Classifier, Decision Tree Classifier).So each algorithm has shown 100% Accuracy then we considered Time Complexity of each algorithm. After comparing each algorithm's Time complexity Decision Tree has shown optimum time. So we predict decision tree as best algorithm.

ADVANTAGES OF PROPOSED SYSTEM

1. By using data mining techniques it takes less time for the prediction of the disease with more accuracy.
2. In this paper we survey different papers in which one or more algorithms of data mining used for the prediction of CKD disease.
3. Applying data mining techniques to CKD disease treatment data can provide as reliable performance as that achieved in diagnosing kidney disease.

5. PROPOSED MACHINE LEARNING ALGORITHMS

The proposed system try to use some well known ML classification algorithms such as Logistic Regression, Decision Trees, Random Forest and Support VectorMachine.Finally based on the performance of each and every individual algorithm, we will decide which algorithm gives best accurate result for identifying the chronic kidney diseases. There are totally 5 modules present in this current application:

1. Import Necessary Libraries
2. Load Dataset Module
3. Data Pre-Processing
4. Train the Model Using Several ML Algorithms
5. Find the Performance of ML Algorithms

1) IMPORT NECESSARY LIBRARIES

In this module initially we need to import all the necessary libraries which are required for building the model. Here we try to use all the libraries which are used to convert the data into meaningful manner. Here the data is divided into numerical values which are easily identified by the system, hence we try to import numpy module and for plotting the data in graphs and charts we used matplot library.

2) LOAD DATASET MODULE

In this module the we try to load the dataset which is downloaded or collected from Kaggle repository. Here we store the dataset names as **ckdisease.zip** file and this dataset contains the following information such as :

3) DATA PRE-PROCESSING MODULE

Here in this section we try to pre-process the input dataset and find out if there are any missing values or in-complete data present in the dataset. If there is any such data present in the dataset, the application will ignore those values and load only valid rows which have all the valid inputs

4) TRAIN THE MODEL USING SEVERAL ML ALGORITHMS

Here we try to train the current model on given dataset using several ML classification algorithms and then try to find out which algorithms suits best in order to identify and classify the input dataset accurately and efficiently. Here we try to use following algorithms on input dataset such as:

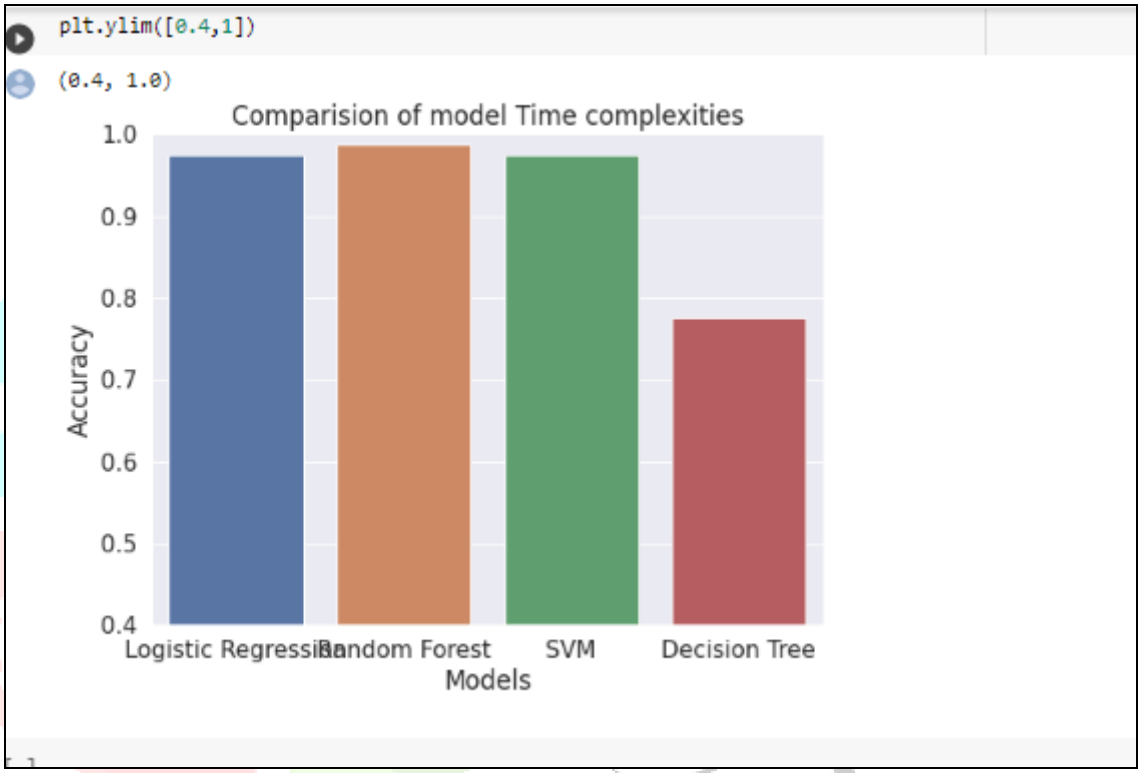
1. Logistic Regression
2. Support Vector Machine
3. Decision Tree
4. Random Forest

5) PERFORMANCE ANALYSIS MODULE

Here in this module we try to compare each and every classification algorithm on given input dataset and then try to find out which one suits best for finding the accurate results. Finally we will identify the best algorithm which give accurate results in very less time. Here we can see **Random Forest** gives more accurate result compared with other ML Algorithms.

6. RESULT AND DISCUSSION

PERFORMANCE ANALYSIS



From the above window we can clearly see Random Forest achieved more accuracy compared with other ML algorithms and hence RF is best for chronic kidney disease prediction for early diagnosis.

7. CONCLUSION

According to the study's findings, the random forest technique and logistic regression can be utilised to better correctly forecast chronic kidney disease. The study found that their precision was 96.25 percent and their accuracy was 97 percent. The accuracy percent of the models utilised in this investigation is significantly greater when compared to earlier research, showing that the models employed in this study are more dependable than those used in previous investigations. The LR technique outperforms the other methods when cross validation measurements are used to forecast chronic kidney disease. But coming to Random forest all the performance metrics are more in order to detect the chronic kidney diseases and hence as a future research could improve on this work by creating a web application that uses these algorithms and a larger dataset than the one used in this study.

8. REFERENCES

1. A. S. Levey, R. Atkins, J. Coresh et al., "Chronic kidney disease as a global public health problem: approaches and initiatives–a position statement from kidney disease improving global outcomes," *Kidney International*, vol. 72, no. 3, pp. 247–259, 2007.
2. V. Jha, G. Garcia-Garcia, K. Iseki et al., "Chronic kidney disease: global dimension and perspectives," *The Lancet*, vol. 382, no. 9888, pp. 260–272, 2013.
3. N. R. Hill, S. T. Fatoba, J. L. Oke et al., "Global prevalence of chronic kidney disease – a systematic review and meta-analysis," *PLoS One*, vol. 11, no. 7, article e0158765, 2016.
4. H. Nasri, "World kidney day 2014; chronic kidney disease and aging: a global health alert," *Iranian Journal of Public Health*, vol. 43, no. 1, pp. 126-127, 2014.
5. G. Abraham, S. Varughese, T. Thandavan et al., "Chronic kidney disease hotspots in developing countries in South Asia," *Clinical Kidney Journal*, vol. 9, no. 1, pp. 135–141, 2016.
6. K. T. Mills, T. Xu, W. Zhang et al., "A systematic analysis of worldwide population-based data on the global burden of chronic kidney disease in 2010," *Kidney International*, vol. 88, no. 5, pp. 950–957, 2015.
7. B. Ene-Iordache, N. Perico, B. Bikbov et al., "Chronic kidney disease and cardiovascular risk in six regions of the world (ISN-KDDC): a cross-sectional study," *The Lancet Global Health*, vol. 4, no. 5, pp. e307–e319, 2016.
8. S. Anand, M. A. Khanam, J. Saquib et al., "High prevalence of chronic kidney disease in a community survey of urban Bangladeshis: a cross-sectional study," *Glob Health*, vol. 10, no. 1, p. 9, 2014.
9. L. Ali, K. Fatema, Z. Abedin et al., "Screening for chronic kidney diseases among an adult population," *Saudi Journal of Kidney Diseases and Transplantation*, vol. 24, no. 3, p. 534, 2013.
10. M. J. Hasan, M. A. Kashem, M. H. Rahman et al., "Prevalence of chronic kidney disease (CKD) and identification of associated risk factors among rural population by mass screening," *Community Based Medical Journal*, vol. 1, pp. 20–26, 2013.
11. M. J. Lysaght, "Maintenance dialysis population dynamics: current trends and long-term implications," *Journal American Soietyc Nephrology*, vol. 13, suppl 1, pp. S37–S40, 2002.
12. M. Bakhshayeshkaram, J. Roozbeh, S. T. Heydari et al., "A population-based study on the prevalence and risk factors of chronic kidney disease in adult population of shiraz, southern Iran," *Galen Medical Journal*, vol. 8, no. 935, p. 935, 2019.