A HYBRID METHOD FOR DETECTION OF KIDNEY DISEASE USING MACHINE LEARNING

Dr. M. Ratnababu, and Dr. M. Raghava Naidu
Professor, Department of ECE, St. Ann's College of Engineering and Technology, Vetapalem Mandal, Chirala, Andhra Pradesh, India
Assistant Professor, Department of CSE, Krishna University College of Engineering and Technology, Machilipatnam, Andhra Pradesh, India

ABSTRACT
The decline of kidney function is known as Chronic Kidney Disease (CKD). Known also as chronic renal disease, chronic kidney disease is an abnormal functioning of the kidneys or a progressive loss of renal function that develops over months or years. is a significant burden on the healthcare system because to the rising patient population, poor prognosis for morbidity and death, and increased risk of developing end-stage renal disease. yearly because of the illness. Therefore, it takes time, the illness's symptoms are not immediately apparent, and a large number of lives are lost before kidney disease is taken seriously in its early stages. Modern Machine Learning (ML) approaches are being used to detect several key health hazards, including the prediction of diabetes, the detection of brain tumors, the identification of COVID-19, the detection of renal illnesses, and many more. Therefore, the illness may be predicted in this study utilizing these machine learning classifiers, hybrid namely KNN and Logistic Regression. Our major goal is to distinguish between different machine learning algorithms based on their correctness in terms of performance. This hybrid method can produce accurate and dependable F1-score and accuracy outcomes.

Keywords: Chronic Kidney Disease, Logistic Regression (LR), Random Forest (RF), K-Nearest Neighbors (KNN).

1.INTRODUCTION
Kidneys are two bean-shaped organs, each about the size of a fist. They are located just below the rib cage, one on each side of the spine. Every day, the kidneys filter about 120 to 150 quarts of blood to produce about 1 to 2 quarts of urine. The key function of the kidneys is to remove waste products and excess fluid from the body through the urine. The production of urine involves highly complex steps of excretion and re-absorption. This process is necessary to maintain a stable balance of body chemicals. The critical regulation of the body 's salt, potassium and acid content is performed by the kidneys and produce hormones that affect the function of other organs [1]. For example, a hormone produced by the kidneys stimulates red blood cell production, regulate blood pressure and control calcium metabolism etc. Chronic kidney disease (CKD) is a major issue worldwide which is a condition characterized by a gradual loss of kidney function over time, 14% of the world population suffer from CKD. Over 2 million people worldwide currently receive treatment with dialysis or a kidney transplant to stay alive, yet this number may only represent 10% of people who need treatment to live. Chronic kidney disease causes more deaths than breast cancer or prostate cancer.
In India, there are approximately one million cases of Chronic Kidney Disease (CKD) every year. It is dangerous to kidney and it produces gradual loss in kidney functionality. Nevertheless, it is unpredictable because its symptoms grow gradually and is not unique to the disorder, it is important to detect CKD at its early stage. Kidneys filter wastes and excess fluids from the blood that are then excreted in excrement. In the early stages of CKD, a few signs or symptoms will be observed [2].
It makes its way as a ground-breaking and actual channel that liberates the body from squander and parlous substances and return supplements, amino acids, insulin, hormones and different basic substances to the circulatory framework. Incidentally things will flip out gravely, however. “Chronic Kidney Disease (CKD) is used at some stage in the world to suggest to any variety of nephritis that returns. “Infection” incorporates any deviation from the urinary organ structure or limit customary, paying very little heed to whether or not it is most likely going to create a man feel unwell or manufacture complexities [3]. It is a typical issue that may influence anybody at any age. It’s assessed that just about 3 million people within the United Kingdom are at risk of CKD.
Chronic diseases pose an important threat to the global health agenda of the 21st century. The rising prevalence of chronic diseases such as chronic kidney disease has serious consequences for health and economic output in developing countries. The rapid increase in common risk factors, especially among the poor, such as diabetes, hypertension, and obesity, would result in even greater and deeper burdens that developing nations are not prepared to cope with. There has been a lack of exposure to chronic diseases, chronic kidney disease in particular, primarily due to the focus of the global health community

on infectious diseases, and lack of consciousness. There is a vital need to concentrate on it and the adoption of more inclusive, cost-effective, and preventive chronic disease strategies by developed countries. There are many hospitals which store the data of chronic kidney disease patients in their database [4]. Through analyzing these data, various patterns can be found that will be helpful for decision-making. Using data mining techniques on these data, it is possible to discover many kinds of knowledge and use this knowledge to predict the disease. There is a massive amount of people who are affected by chronic kidney disease. In Bangladesh maximum are not concerned about the disease. As a result, the affected people by the disease are increasing day by day. If people could detect or predict if they are affected or going to be affected then it could be controlled. People could take the necessary steps to not be affected. For predicting the disease, predictive can be used. There are several methods that can be used like, classification, regression, categorization. Among those, many people think classification is the best. After accomplishing the research, it will help to predict Chronic Kidney disease. People will be aware of the disease and their health condition [5]. The main objectives of the research are to predict the disease using a machine learning algorithm, warning whether a person at risk of the disease or not.

In diagnosis of CKD predicting results of patients remains an important for researchers, individuals, medical and health care systems. An early detection of CKD can improve the quality of life to a greater extent [6]. For this purpose, machine learning algorithms performed an important rule and applied it to predict chronic kidney diseases. Hence in this analysis a novel approach to optimize prediction method for chronic kidney disease is presented.

2. LITERATURE SURVEY

S.Revathy, B.Bharathi, P.Jeyanthi, M.Ramesh et. al., [7] presented CKD Prediction using ML Models. This analysis uses data preprocessing, data transformation and various classifiers to predict CKD and also proposes best Prediction framework for CKD. The results of the framework show promising results of better prediction at an early stage of CKD. Devika R, Sai Vaishnavi Avilala, V. Subramaniyaswamy et. al., [8] presents a Comparative Study of Classifier for CKD prediction using Naive Bayes, KNN and Random Forest. They compared the overall performance of the used classifiers with other current classifiers. The outcome after conducted research is that the performance of RF classifier is finest than NB and KNN. V Ganapathi Raju, K Prasanna Lakshmi, K. Gayathri Praharshitha, Chittampalli Likhitha et. al,[9] presented Prediction of CKD using Data Science. This research work is primarily concentrated on finding the best suitable classification algorithm which can be used for the diagnosis of CKD based on the classification report and performance factors. Zixian Wang, Jae Won Chung, Xilin Jiang, Yantong Cui, Muning Wang, Anqi Zheng et. al., [10] presents ML based prediction system for CKD Using Associative Classification Techniques. This study analyzes chronic kidney disease using machine learning techniques based on a Chronic Kidney Disease (CKD) dataset from the UCI machine learning data warehouse. CKD is detected using the Apriori association technique for 400 instances of chronic kidney patients with 10-fold-cross-validation testing, and the results are compared. David Y. Gaitonde, David L. Cook, Md; And Ian M. Rivera, et. al., [11] presents a Chronic Kidney Disease: Detection and Evaluation. A multidisciplinary approach between primary care physicians, nephrologists, and other subspecialists for implementing early interventions, providing education, and planning for advanced renal disease is key for effective management. Asif Salekin and John stankovic [12] used novel approach. They got findings on a dataset consisting of 400 records and 25 attributes resulting in a patient prone to CKD or not. In order to achieve results, they used KNN, random forest and Neural Network algorithms. They used wrapper methodology for feature reduction which finds CKD with high accuracy. Sharma et. al. [13] evaluated different classification methods for diagnosis of Chronic Kidney Disease (CKD). The dataset used for this analysis consisted of 400 instances and 24 attributes. The authors tested twelve classification techniques by applying them to the CKD results. The predictor parameters used for performance measurement are prognostic accuracy, precision, sensitivity and specificity. Sahil Sharma, Vinod Sharma, and Atul Sharma [14], tested 12 different classification algorithms on various datasets by each 400 records and 24 attributes. They compared their expected outcomes with actual results in order to determine predictive accuracy. They used metrics such as precision, sensitivity, accuracy and specificity for measuring the performance of the classifiers. Note that Chronic Kidney Disease (CKD) is not uncommon. Radha et. al. [15] the classification algorithms were expected to be taken into consideration for the diagnosis of Chronic Kidney Disease. On entirely different algorithms such as Naive Bayes, Decision Tree and K-Nearest Neighbor the expected results were carried out. The expected outcome suggests that the KNearest Neighbour algorithm delivers a better outcome than the

other classification algorithms.

3. PREDICTION OF CHRONIC KIDNEY DISEASE

In this section A novel approach to optimize prediction method for chronic kidney disease using machine learning is presented. The block diagram of presented approach is shown in Fig. 1.

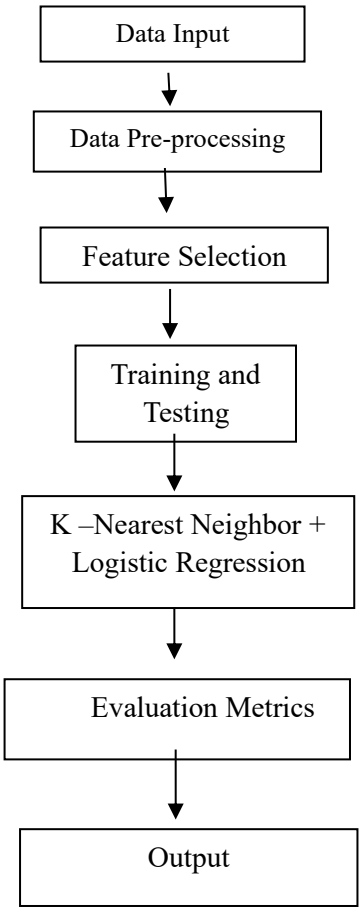

Figure: 1 Block Diagram of Chronic Kidney Disease Prediction

In this analysis, 400 chronic kidney disease datasets are used from the UCI machine learning repository. In this dataset, there are 25 attributes, 24 are predictive variables and 1 attribute is decision class. In predictive 24 attributes some attributes are nominal some attributes are numerical. Therefore convert nominal attributes into numerical attributes using a mapping function. This dataset contains the nominal values and convert this numeric value using a mapping function. And now our data set is full of numeric value. It pre-processed the unrefined medical data by removing the missing values to enhance prediction capabilities. It also conducted data-transformation to make them useful for the machine learning models, which are limited to process non numerical data. The non-numerical data in the dataset are in the form of 'present', 'not present', 'normal', 'abnormal', 'yes', 'no', 'good', and 'poor'. The non-numerical data are identified and transformed into numbers. The 'normal', 'present', 'yes', and 'good' values for nominal attributes are replaced by '1' and 'abnormal', 'not present', 'no', and 'poor' values are replaced by '0'. Feature selection is the process of reducing the number of input variables when developing a predictive model. It is desirable to reduce the number of input variables to both reduce the computational cost of modeling and, in some cases, to improve the performance of the model. Feature selection methods can be used in data pre-processing to achieve efficient data reduction. This is useful for finding accurate data models.

Data Pre-Processing is that stage where the data that is distorted, or encoded is brought to such a state that the machine can easily analyze it. A dataset can be observed as a group of data objects. Data objects are labeled by a number of features that ensures the basic features of an object, such as the mass of a physical object or the time at

which an event ensured. In the dataset there may be missing values, they can either eliminated or estimated. The most common method of dealing with missing values is filling them in with mean, median or mode value of respective feature. After the fruitful completion of the data Pre-processing strategy, all features of data are used for selection. To reduce dimensionality, the process of attribute selection has been attached. Small subsets of relevant features are taken from the presented dataset to obtain improved prediction rates. The classified model particularly figures out and takes all inputs for the kidney disease dataset. Four different machine learning approaches which are Logistic Regression, Decision Tree Random Forest, and K-Nearest Neighbors are to be implemented in this dataset.

Three distinct classifiers receive extracted Features that have been tagged with the appropriate diseases for training as well as testing. This study is performed using three machine learning algorithms to evaluate the CKD or NOT CKD performance. The algorithms are: KNN, Random Forest and Logistic Regression (LR). The dataset that is taken is divided into two groups, one for testing the samples and another for training the samples. The ratio for testing and training data is 30% and 70% respectively.These models are evaluated using the testing set.

The machine learning algorithms used to classify CKD are Logistic Regression, Random Forest and K-Nearest Neighbor. Logistic regression: Logistics regression is a mathematical model that, while there are many more complicated extensions, uses logistical functions in its basic form to model a binary dependent variable. Regression analysis measures logistics model parameters by logistic regression (a form of binary regression). The statistical Binary Logistic Model has a dependent variable of two possible values, for example pass/fail, suggesting the two values ”0” or ”1”.

K-Nearest Neighbor: The K-Nearest Neighbor is known as one of the most used algorithms in machine learning due to its versatility. Moreover, like other algorithms, the learning stage is not required. In 1970, KNN was used for statistical estimates and pattern recognition. In data mining, K-Neighborhood is called a classified algorithm and a lazy algorithm. For mathematics, the range between two points is known to be space. There are several matrices within the distance. Among them, the distance from Euclidean is counted as the distance from the universe.

Here in this analysis, the accuracy, Recall and F1 score evaluation metrics are used. By using these evaluation metrics the outcomes are predicted.

4. RESULT ANALYSIS

In this section, performance of a Novel Approach To Optimize Prediction Method For Chronic Kidney Disease Using Machine Learning is discussed.

Table: 1 Performance Comparison Table

ML models	**Accuracy**	**F1 score**
Hybrid (LR+ KNN)	88	95
RF	80	91

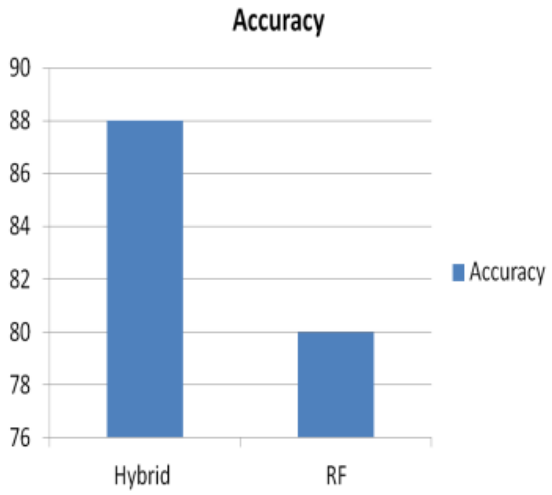

Fig. 2: Accuracy Comparison Graph

In Fig.2 accuracy comparison graph is observed between hybrid and RF.

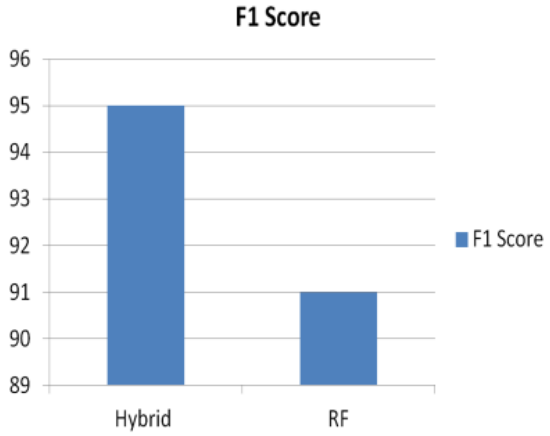

Fig. 3: F1-Score Comparison Graph

In Fig.3 F1 Score comparison graph is observed between hybrid and RF.

4. CONCLUSION

In this examination a unique technique to optimize the prediction algorithm for chronic kidney disease by utilizing machine learning. Here, machine learning is applied to create a fresh approach method employing hybrid method. This method trains and tests the proposed model using data from the UCL ML repository. The chronic kidney disease is reliably predicted by the new ML as either CKD or Non-CKD. The accuracy, recall, and F1 score of the model performance that is being provided are assessed. In comparison to hybrid and RF ML classifiers, the outcomes of our innovative method RF ML model are more accurate and superior. The hybrid model achieved better results interms of accuracy and F1 Score.

REFERENCES

[1] S. Vijaiyanti and s. Dayanand, "Kidney Disease Prediction Using SVM and ANN Algorithms", vol. 6, no. 2, March 2015, ISSN 22296166.
[2] Mahfuzah Mustafa, Mohd Nasir Tab et al., "Comparison between KNN and ANN Classification in Brain Balancing Application via Spectrogram Image", *Journal of Computer Science & Computational Mathematics*, vol. 2, no. 4, pp. 17-22, April 2012.
[3] K Ross, K Leung, Ying Wang et al., "Using a multi-staged strategy based on machine learning and mathematical modelling to predict genotype-phenotype risk patterns in diabetic kidney disease: a prospective case-control cohort analysis", *BMC-Nephrology*, pp. 1-9, 2013.
[4] ZarPhyu T, Oo NN. Performance comparison of feature selection methods. MATEC Web Conf. 2016;42:2–5.
[5] Rady EA, Anwar AS. Informatics in Medicine Unlocked Prediction of kidney disease stages using data mining algorithms. Informatics Med. 2019;15(2018):100178.
[6] Salekin A, Stankovic J. Detection of Chronic Kidney Disease and Selecting Important Predictive Attributes. In: Proc. - 2016 IEEE Int. Conf. Healthc. Informatics, ICHI 2016, pp. 262–270, 2016.
[7] S. Revathy, B. Bharathi, P. Jeyanthi, M. Ramesh, "Chronic Kidney Disease Prediction using Machine Learning Models", International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-9 Issue-1, October 2019
[8] Devika R, Sai Vaishnavi Avilala, V. Subramaniyaswamy, "Comparative Study of Classifier for Chronic Kidney Disease prediction using Naive Bayes, KNN and Random Forest", Proceedings of the Third International Conference on Computing Methodologies and Communication (ICCMC 2019), IEEE Xplore Part Number: CFP19K25-ART; ISBN: 978-1-5386-7808-4
[9] V Ganapathi Raju, K Prasanna Lakshmi, K. Gayathri Praharshitha, Chittampalli Likhitha, "Prediction of chronic kidney disease (CKD) using Data Science", Proceedings of the International Conference on Intelligent Computing and Control Systems (ICICCS 2019), IEEE Xplore Part Number: CFP19K34-ART; ISBN: 978-1-5386-8113-8

[10] Zixian Wang, Jae Won Chung, Xilin Jiang, Yantong Cui, Muning Wang, Anqi Zheng, " Machine Learning based prediction system for Chronic Kidney disease Using Associative Classification Techniques", International Journal of Engineering and Technology, 7 (4.36) (2018) 1161-1167

[11] David Y. Gaitonde, David L. Cook, Md; And Ian M. Rivera, "Chronic Kidney Disease: Detection and Evaluation", 2017 American Academy of family physicians, doi: afp/2017/1215/p776-s1. html

[12] Asif Salekin, John Stankovic, "Detection of Chronic Kidney Disease and Selectiing Important Predictive Attributes," Proc. IEEE International Conference on Healthcare Informatics (ICHI), IEEE, Oct. 2016, doi:10.1109/ICHI.2016.36.

[13] Sharma, S., Sharma, V., Sharma, A. (2016). Performance based evaluation of various machine learning classification techniques for chronic kidney disease diagnosis. arXiv preprint arXiv:1606.09581.

[14] Sahil Sharma, Vinod Sharma, Atul Sharma, "Performance Based Evaluation of Various Machine Learning Classification Techniques for Chronic Kidney Disease Diagnosis," July18, 2016.

[15] Radha, N., Ramya, S. (2015). Performance analysis of machine learning algorithms for predicting chronic kidney disease. Int. J. Comput. Sci. Eng. Open Access, 3, 72-76