ISSN 2063-5346

EFFECTIVE APPROACH TO PREDICT CHRONIC KIDNEY DISEASE USING MACHINE LEARNING

[1]Dr. Rajesh KS, [2]Rashmi N Kuppast, [3]Dr. Kalpana D, [4]Dr.Mahantesh Mathapati, [5]Srinivas R

Article History: Received: 10.05.2023 Revised: 29.05.2023 Accepted: 09.06.2023

Abstract

— People now commonly suffer from chronic kidney disease (CKD). By detecting and treating those who are at risk for this condition as soon as feasible, a variety of serious problems, such as end-stage renal disease, elevated risk, and cardiovascular disease, may be prevented. Medical researchers can get a lot of help from the machine learning algorithm in accurately diagnosing the disease at the very beginning. Algorithms for machine learning and Big Data platforms have recently been combined to improve healthcare. This work presents hybrid machine learning methods that integrate extraction of the feature strategies and various algorithms of machine learning under classification technique related to massive data platforms to identify chronic kidney disease (CKD). In this study, logistic regression (LR), random forest (RF), decision tree (DT), support vector machine (SVM), Naive Bayes (NB), and gradientboosted trees were employed as six ensemble learning strategies for machine learning classification tasks (GBT Classifier). The results were validated using four evaluation techniques: accuracy, precision, recall, and F1-measure. The results demonstrated that the chosen features had helped SVM, DT, and GBT Classifiers operate at their peak levels.

Keywords— Chronic kidney, Naive Bayes (NB), decision tree (DT), logistic regression (LR), Gradient-Boosted Trees (GBT Classifier) and Random Forest (RF).

[1,2,4] Dept Artificial Intelligence &Machine Learning, RRCE,

[3]Prof and Head,Dept. of Prosthodontics Dayananda Sagar College of Dental Sciences and Hospital,

[5]Dept of CSE,AIEMS

DOI:10.48047/ecb/2023.12.9.93

I. INTRODUCTION

In terms of the current state of society's health, this Chronic Kidney Disease (CKD) is viewed as a serious hazard. Regular laboratory testing can identify chronic kidney disease, and there are therapies available that can delay the onset of the condition, stop it from progressing, lessen the risk of cardiovascular disease and its complications, as well as enhance survival and quality of life. Lack of water intake, smoking, a poor diet, insufficient sleep, and many other factors can lead to CKD. In 2016, this condition afflicted 753 million people worldwide, 417 million of them were female and 336 million were male. The majority of the time, the illness is discovered at its advanced form, which can occasionally result in renal failure. The International Society of Nephrology's Global Burden Disease 2010 report found that chronic kidney disease (CKD) has become a major cause of mortality globally, with the number of fatalities rising by 82.3% during the previous two decades [1, 2]. Additionally, as more patients develop end-stage renal disease (ESRD), dialysis or kidney transplants are needed to save their lives [1,3,4].

A significant problem for the medical profession is the early detection and treatment of CKD. The treating physician (nephrologist) is called upon to treat the aforementioned systemic signs in addition to slowing the disease's progression to more advanced stages and, if possible, suspending it. ML techniques have become essential tools for a variety of applications in the health sector, including the early identification of some chronic conditions, thanks to the growth of communication technology, sensor networks, data science, and statistical processing. A machine learning-based strategy for the CKD disease will be given in the current study endeavour. The following are the primary benefits of the methods used:

- A phase in data pre-processing that makes use of the Synthetic Minority Oversampling Technique (SMOTE) is essential to assure that the dataset instances are detached regularly, as a result, it creates efficient classification models to forecast the risk for the development of CKD.
- This consists of three distinct sub-steps: (i) statistical description of numerical attributes, (ii) measurement of relevance using three separate approaches, and (iii) tabulation of frequency of occurrence for nominal features.
- The performance of several models is compared and evaluated using the recent used metrics like Precision, Recall F-Measure, and Accuracy. Two kidneys, which are essential organs for the body's healthy operation, are placed in the peritoneal cavity in the rear of the human body. The primary job of the kidneys is to maintain a healthy equilibrium of water, ions, acids, calcium, phosphorus, magnesium, potassium, and other trace components in the body. At the same time the kidneys discharge hormones like erythropoietin, vitamin D, and rennin. Erythropoietin primarily promotes the growth and maturation of red blood cells in the bone marrow, whereas vitamin D controls the body's levels of calcium and phosphorus, as well as bone structure and many other processes. Additionally, hormones that control blood pressure, bone metabolism, fluid balance and vascular calcifications work through the kidneys. Lastly, the kidneys flush out all waste products of metabolism, medications, and other pollutants that enter the body. The main conclusion of this investigation can be inferred from a performance evaluation in which all models showed extraordinarily high performance, with Rotation Forest earning the greatest results across all criteria.

The additional work is divided into the following section. This related works that

use ML to exploit the CKD state in section 2. In addition, the dataset and examine the chosen methodology in Section 3. Section 4 discuss about the how to evaluate the result using evaluation matrix on different algorithm. Section 5 concludes the essay and lays out the paper's future course.

II. LITERATURE SURVEY

J. Snegha et al. [5] suggested a system that makes use of back propagation neural networks and the Random Forest algorithm, among other data mining approaches. Here, they evaluate the two algorithms and discover that the Back Propagation model produces greatest results since it makes use of the Feedforward Neural Network, a supervised learning network.

Mohammed et al.[6] claims that a strategy for CKD makes use of ACO and density-based feature selection. Through wrapper methods, the system chooses features. The creation of a CKD prediction system utilize ML model like as Naive Bayes K-Nearest Neighbor, Support Vector Machine, Regression, Multi-Layer Perceptron Algorithm, Decision Tree, Random Forest, and was proposed by Baisakhi Chakraborty et al.[7]in 2019. These are employed, and the potency of each is assessed in light of the precision, accuracy, and recall results. Random Forest is finally used as it gets the better result compared to all other algorithm. Arif-Ul-Islam et al.[8] proposed a method to predict illness using J48 Decision Tree, Ant-Miner, and Boosting Classifiers. The two objectives in this study, are to evaluate boosting algorithms' efficacy in CKD detection and to create rules that illustrate relationships among the various CKD characteristics. Findings from the trial showed that AdaBoost performed marginally worse than Logit Boost. For the purpose of predicting CKD, S.Belina et al.[9] presented a system utilising an extreme learning machine and ACO has been presented. For classification, a MATLAB tool is utilised, and ELM has certain optimization restrictions. This approach enhances the SLFNs' sigmoid additive type. SiddheshwarTekale et al.[10] described a machine learning system that makes use of SVM, Decision tree algorithms. By contrasting the 2 methods, it was determined that SVM produces the best results. In order for clinicians to examine patients more quickly, its prediction procedure requires less time. The Back Propagation model of Neural Network technique for prediction was reported by Nilesh Borisagar et al.[11]. Algorithms discussed here are Bayesian regularization, Levenberg, Scaled Conjugate, and the robust back propagation algorithm. For the purposes of implementation, Matlab R2013a is employed. Scaled merge gradient and back propagation to get more effective results than Levenberg and Bayesian regularization in terms of training time.

Guneet Kaur et al.[12] presented a method for forecasting CKD using Hadoop's data mining algorithms. They employ two KNN and SVM-based data mining classifiers. Here, the manually chosen data columns are used to do the prediction analysis. In this system, SVM classifier provides better accuracy than KNN.

A system that analyses a patient's kidney disease and computes the findings automatically using the patient's data set was proposed by Neha Sharma[13] et al. In this case, rule-based prediction is applied. The neuro-fuzzy approach is used by this system, and the results were calculated mathematically.

Kai-Cheng Hu et al.[14] suggested a system for clustering that makes use of a multiple pheromone table based on ACO. Here, depending on the qualities of the patterns, the challenge was separated into a number of distinct patterns. Here, two pheromone tables are employed, one to track promising information and the other to store details of unpromising information, increasing the likelihood that someone will search for instructions.

Data mining techniques and conversions were used to preprocess data by Basma Boukenze et al. [15] to learn more about the relationship between measurement parameters and patient survival. Two data mining algorithms were utilised to extract knowledge, create decision rules, and forecast patient survival. They discussed the importance of employing data mining to explore key parameters. With the use of dialysis information gathered from four separate sites, their novel idea was put into practice and tested. Additionally, their approach lowered the expense and work involved in choosing participants for clinical studies. Based on expected outcomes and significant parameters discovered during their study, the patients were chosen.

Veenita Kunwar et al.[16] used an artificial neural network (ANN) and naïve Bayesian classification to predict chronic kidney disease (CKD) in their study. Their findings demonstrated that naïve Bayesian outperformed artificial neural networks in terms of accuracy. Additionally, it was noted that a lot of CKDs were investigated and identified using classification techniques.

Data mining techniques have been successfully applied in medical applications, as shown by Swathi Baby P et al [17]. Data from kidney disease patients were gathered for their investigation. The outcomes demonstrated the versatility of data mining in a range of medical applications.

Microsoft Azore was employed in the study conducted by W. Gunarathne, K. et al.[18] to forecast the patient status of CKD. They tested four distinct algorithms—Multiclass Decision Forest, Multiclass Decision Jungle, Multiclass Decision Regression, and Multiclass Neural Network—by taking into account 14 of the available 25 criteria. They discovered that Multiclass Decision Forest had 99.1% accuracy, which was the best performance.

P. Yildirim [19] investigated how sampling algorithms could be used to foretell chronic kidney disease. The three sampling algorithms, Resample, SMOTE, and Spread Sup Sample, were compared to see how they affected the multilayer perceptron classification algorithm's prediction. The study shown that sampling algorithms may enhance the performance of the classification algorithm, and among the sampling algorithms, the resample approach has a greater accuracy.

In order to predict the course of CKD, J. Xiao et al. [20] developed and compared nine machine learning models, including LR, Elastic Net, ridge regression, lasso regression, SVM, RF, XGBoost, and neural networks. They used the 551 CKD follow-up patients' available clinical characteristics. They come to the conclusion that linear models, with an average AUC above 0.87 and accuracy above 0.8 and 0.8, respectively, have the greatest overall predictive potential.

III.DATA SET AND METHODOLOGY

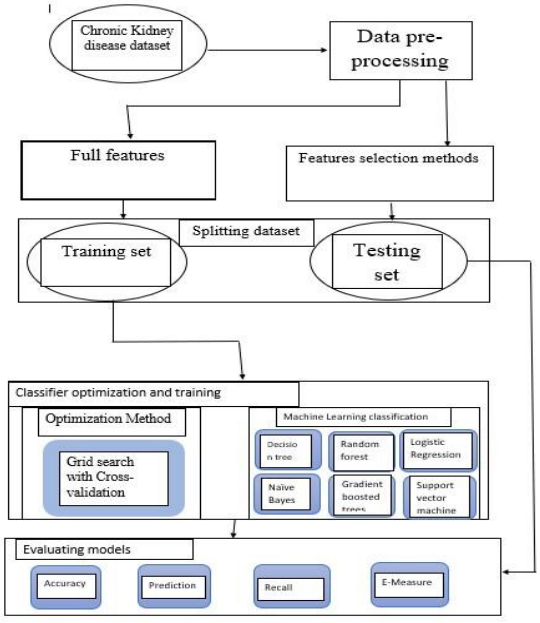

Fig1: Architecture of Chronic Kidney Disease

Figure 1 illustrates the two basic approaches that make up the proposed system for forecasting chronic kidney disease. The first method chooses the most important features

from the datasets on chronic kidney disease using feature selection methods. The second method uses ensemble learning, DT, RF, SVM, LR, NB, and ML algorithms on this whole set of data and the selected features to predict CKD. Data collection is the first stage in the suggested strategy, which involves six steps and uses the dataset CKD which is from UCI machine learning library. Null values are handled at the data preprocessing stage, which is the second step. The third stage will involve choosing the key features using feature techniques. The parameters of ML and ensemble learning algorithms are optimized in the fourth stage make use of stratified cross-validation. The following subsections contain a detailed explanation of each stage.

Dataset

From the UCI repository dataset of CKD are utilizing in this instance. 400 samples from two separate classes are included. There are 25 qualities total, 11 of which are numerical, 13 nominal, and 1 class attribute. There are several missing values in the data set. Here, the dataset holds the data about the patient such as blood pressure, age, sugar, RBC count, specific gravity, albumin etc.,

High blood pressure and diabetes can contribute to CKD. Our numerous organs are impacted by diabetes, and excessive level of diabetes will follow. Therefore, it is crucial to find the illness as soon as feasible. In order to anticipate the sickness, this study improvises a number of machine learning approaches.

Pre-Processing

The process of converting distorted or encoded data into a format that can be quickly and easily analysed by a machine is known as data pre-processing. A collection of data elements can be seen as a dataset. The attributes that are considered as an object, like the bunch of a physical object or the precise moment where we can see the event has occurred, are secured by a variety of qualities that are used to identify data items. Missing values in the dataset may need to be approximated or deleted. Filling in missing values with the mean, median, or mode value of the corresponding feature is the most typical approach of handling missing data. We must change object-typed numerical values to float64 types since object values cannot be used for analysis. The attribute value currently appearing most frequently in that column attribute is utilized to replace the null values for the category attributes. By assigning each distinct attribute value to an integer, label encoding converts category data into numeric attributes. This converts the characteristics to an int type automatically. Every missing value in that attribute column is replaced by the mean value, which is calculated beforehand from each column. For this function, which calculates the mean value for each column, we're utilizing a function called imputer. When the replacement and encoding are finished, The data needs to be evaluated, validated, and trained. The process of actually training our algorithms to create a model occurs during the training of the data phase. The dataset's validation section is used to check the precision of our multiple model fits or to improve the model. Our model's premise is tested using the data.

Feature Selection

The ability to identify the crucial features in a dataset is one of the key advantages of feature selection algorithms. Different classification techniques are used to get better results and the model's execution time is decreased with the help of proper feature selection made. The chi-squared feature selection and Relief-F methods accustomed to identify the group of significant characteristics taken from a database. acronyms and abbreviations

Splitting the Dataset

For training, 80% of the CKD datasets are used, and for testing, 20%. Utilizing stratified cross-validation, the models were trained and optimized, and the cross-validation outcomes were documented. Results from the testing set, which we utilized to assess the models, have been recorded.

Models' Optimization and Training

Optimization Methods

The models are improved and the hyper parameters are tuned by using grid search with K-Fold cross-validation. The usually used technique for tuning parameter optimization is grid search. Each tuning parameter should first have a collection of principles established by users. Then model makes a choice of the hyper parameter with the highest performance after evaluating each possible value for the hyper parameters.

The dataset is partitioned into k folds of equal size for K-Fold cross validation. Hence, the classifiers are tested in the remaining time after the k-1 group training has been completed. This procedure is performed after the 10 folds have each been presented used as a test set. The efficacy of the classifiers for each k is also assessed. Finally, the evaluation classifier is developed based on average performance.

Machine Learning Models

Here is a list of the categorization models that were utilized in the study:

• Decision tree (DT): For classification problems including a well-liked target variable, it could be a supervised learning strategy. Each distinct and continuous input and output variable is supported by a decision tree. This approach applies decision trees to any classification and regression problem when the sample or population is split into two or more groups that are exactly the same, called subpopulations, to support the main splitter of the input parameter.

• Random forest (RF): It is a technique for supervised computer learning. Simply defined, it gathers a lot of trees and combines them for more precise prediction [21][22].Problems with binary categorization were tackled using logistic regression (LR). In order to predict the chance of different labels for an unlabeled observation, LR employs a logistic or sigmoid function [23].

• Support Vector Machines: A popular supervised learning model for classification issues is Support Vector Machines (SVM) [24]. The goal of the SVM method is to determine the ideal hyperplane that, with the widest margins between these two classes, optimally separates all objects of one class from those of the other. To achieve enough computing efficiency, objects that are far from the boundary are excluded from the calculation, while other data points that are close to the boundary are kept and designated as "support vectors" [24]. Different kernel functions for the SVM method include the radial basis function (RBF), linear, sigmoid, and polynomial. Based on the outcomes of the nested cross-validation, the radial basis function has been selected for this investigation [25,26].

• Naive Bayes (NB): Using the Bayes theorem, the Nave Bayes approach trains a classifier. In other words, the Naive Bayes method was used to train a probabilistic classifier. It creates a probability distribution for a certain observation over a number of classes.

Gradient-Boosted Trees (GBTs): Another method for training a collection of decision trees is to utilize the algorithm known as Gradient-Boosted Trees (GBTs). Every decision tree, however, undergoes sequential training. In order to optimize each new tree, this method uses data from previously trained trees. Consequently, with each new tree, the model gets better. The training of a model could take longer with GBT since it trains one tree at a time. When

a lot of trees are used in an ensemble, overfitting is also a possibility. However, a GBT ensemble may have shallow trees, which facilitates training. Gradient boosting is a method for continual training of many decision trees. This model predicts the label for unit training sample using the current ensemble and compares the forecast to the actual label for each iteration.

IV EVALUATION MATRIX

Equations 1-4 list four standard metrics that are used to evaluate the models: True positive, true negative, false positive, and false negative are all abbreviated as TP, TN, FP, and FN, respectively. Other indicators include accuracy, precision, recall, and F1-score.

The predicated output cases from dataset based on exact ratio

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad \text{.....1}$$

Precision value where predicted, the ratio of all patients with CKD (true positive and false positive).

$$Precision = \frac{TP}{TP + FP} \quad \text{.....2}$$

The actual number ratio of CKD patients are properly recognized to the total no of patients.

$$Recall = \frac{TP}{TP + FP} \quad \text{.....3}$$

The accuracy of the test is the harmonic mean between precision and recall.

$$F1 = \frac{2 - precision - recall}{precision - recall} \quad \text{.....4}$$

Support: It is the exact number of consequence that is present in each class of the predicted consequence.

Result of CKD predicting using SVM

	Precision	Recall	F1 score	support
CKD	1.00	0.94	0.97	83
Not CKD	0.88	1.00	0.94	37
Accuracy			0.96	120

V. CONCLUSION

A model that aids in the prediction of chronic kidney disease is suggested as a result. Low-cost early CKD detection and management can relieve the burden, enhance diabetes and cardiovascular disease (including hypertension) outcomes, and considerably cut patient morbidity and death. Our goal is to suggest a simple however effective approach that enables both patients and doctors to predict chronic kidney disease at an early stage. A computer-assisted diagnostic system can be used by doctors and radiologists to assist in reaching more accurate diagnostic results. Doctors may see more patients in less time using this method. The amount of features needed by the prediction algorithm can be decreased with the use of effective feature selection techniques, which also reduces the number of medical tests necessary. For better CKD prediction, Future research should investigate different supervised and unsupervised machine learning techniques, as well as feature selection techniques with additional performance metrics . To compile the most recent information on CKD diagnosis from diverse global locations. The reliability of this may be impacted by the projected small sample size (400 instances). As a result, for higher accuracy, the dataset size has to be expanded in the future.

REFERENCES

1. J. RADHAKRISHNAN ET AL, "TAMING THE CHRONIC KIDNEY DISEASE EPIDEMIC: A GLOBAL VIEW OF SURVEILLANCE EFFORTS," KIDNEY INT., VOL. 86, (2), PP. 246-250, 2014.
2. R. LOZANO ET AL, "GLOBAL AND REGIONAL MORTALITY FROM 235 CAUSES OF DEATH FOR 20 AGE GROUPS IN 1990 AND 2010: A SYSTEMATIC ANALYSIS FOR THE GLOBAL BURDEN OF DISEASE STUDY 2010," THE LANCET, VOL. 380, (9859), PP. 2095-2128, 2012.
3. R. RUIZ-ARENAS ET AL, "A SUMMARY OF WORLDWIDE NATIONAL ACTIVITIES IN CHRONIC KIDNEY DISEASE (CKD) TESTING," EJIFCC, VOL. 28, (4), PP. 302, 2017.
4. Q. ZHANG AND D. ROTHENBACHER, "PREVALENCE OF CHRONIC KIDNEY DISEASE IN POPULATION-BASED STUDIES: SYSTEMATIC REVIEW," BMC PUBLIC HEALTH, VOL. 8, (1), PP. 117, 2008.
5. J. SNEGHA, "CHRONIC KIDNEY DISEASE PREDICTION USING DATA MINING", INTERNATIONAL CONFERENCE ON EMERGING TRENDS, 2020.
6. JAMALUDDIN J, MOHAMED YASSIN MS, JAMIL SN, MOHAMED KAMEL MA, YUSOF MY. A CLINICAL AUDIT OF THE DIAGNOSIS AND MANAGEMENT OF CHRONIC KIDNEY DISEASE IN A PRIMARY CARE CLINIC. MALAYS FAM PHYSICIAN. 2021.
7. BAISAKHI CHAKRABORTY, "DEVELOPMENT OF CHRONIC KIDNEY DISEASE PREDICTION USING MACHINE LEARNING", INTERNATIONAL CONFERENCE ON INTELLIGENT DATA COMMUNICATION TECHNOLOGIES, 2019.
8. ARIF-UL-ISLAM AND SHAMIM RIPON. "RULE INDUCTION AND PREDICTION OF CHRONIC KIDNEY DISEASE USING BOOSTING CLASSIFIERS, ANT-MINER AND J48 DECISION TREE." 2019 NTERNATIONAL CONFERENCE ON ELECTRICAL, COMPUTER AND COMMUNICATION ENGINEERING (ECCE) (2019): 1-6.
9. V.J SARA, S.BELINA AND KALAISELVI, K, ANT COLONY OPTIMIZATION (ACO) BASED FEATURE SELECTION AND EXTREME LEARNING MACHINE (ELM) FOR CHRONIC KIDNEY DISEASE DETECTION (2018). INTERNATIONAL JOURNAL OF ADVANCED STUDIES OF SCIENTIFIC RESEARCH, VOL. 4, NO 1, 2019
10. SIDDESHWAR TEKALE, "PREDICTION OF CHRONIC KIDNEY DISEASE USING MACHINE LEARNING, INTERNATIONA JOURNAL OF ADVANCED RESEARCH IN COMPUTER AND COMMUNICATION ENGINEERING, 2018
11. BORISAGAR, N., BARAD, D., RAVAL, P. (2017). CHRONIC KIDNEY DISEASE PREDICTION USING BACK PROPAGATION NEURAL NETWORK ALGORITHM. IN: MODI, N., VERMA, P., TRIVEDI, B. (EDS) PROCEEDINGS OF INTERNATIONAL CONFERENCE ON COMMUNICATION AND NETWORKS.
12. GUNEET KAUR, "PREDICT CHRONIC KIDNEY DISEASE USING DATA MINING IN HADOOP, INTERNATIONAL CONFERENCE ON INVENTIVE COMPUTING AND INFORMATICS, 2017.
13. NEHA SHARMA, ER. ROHIT KUMAR VERMA, "PREDICTION OF KIDNEY DISEASE BY USING DATA MININING TECHNIQUES", PREDICTION OF KIDNEY DISEASE BY USING DATA MININING TECHNIQUES, (2016)
14. KAI-CHENG HU, "MULTIPLE PHEROMONE TABLE BASED ON ANT COLONY OPTIMIZATION FOR CLUSTERING", HINDAWI, RESEARCH ARTICLE, 2015.
15. BASMA BOUKENZE, HAJAR MOUSANNIF AND ABDELKRIM HAQIQ, "PERFORMANCE OF DATA MINING TECHNIQUES TO PREDICT IN HEALTHCARE CASE STUDY: CHRONIC KIDNEY FAILURE DISEASE", INTERNATIONAL JOURNAL OF DATABASE

MANAGEMENT SYSTEMS (IJDMS), (2016)

16. VEENITA KUNWAR, KHUSHBOO CHANDEL, A. SAI SABITHA, AND ABHAY BANSAL, "CHRONIC KIDNEY DISEASE ANALYSIS USING DATA MINING CLASSIFICATION TECHNIQUES", IEEE, (2016).
17. SWATHI BABY P AND PANDURANGA VITAL T, "STATISTICAL ANALYSIS AND PREDICTING KIDNEY DISEASES USING MACHINE LEARNING ALGORITHMS", INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) (2015)
18. W. GUNARATHNE, K. PERERA AND K. KAHANDAWAARACHCHI, "PERFORMANCE EVALUATION ON MACHINE LEARNING CLASSIFICATION TECHNIQUES FOR DISEASE CLASSIFICATION AND FORECASTING THROUGH DATA ANALYTICS FOR CHRONIC KIDNEY DISEASE (CKD)," IN BIOINFORMATICS AND BIOENGINEERING (BIBE), 2017 IEEE 17TH INTERNATIONAL CONFERENCE ON, 2017.
19. P. YILDIRIM, "CHRONIC KIDNEY DISEASE PREDICTION ON IMBALANCED DATA BY MULTILAYER PERCEPTRON: CHRONIC KIDNEY DISEASE PREDICTION," IN COMPUTER SOFTWARE AND APPLICATIONS CONFERENCE (COMPSAC), 2017 IEEE 41ST ANNUAL, 2017.
20. J. XIAO ET AL, "COMPARISON AND DEVELOPMENT OF MACHINE LEARNING TOOLS IN THE PREDICTION OF CHRONIC KIDNEY DISEASE PROGRESSION," JOURNAL OF TRANSLATIONAL MEDICINE, VOL. 17, (1), PP. 119, 2019.
21. L. BREIMAN, "RANDOM FORESTS," MACH. LEARNING, VOL 45, (1), PP. 5-32, 2001.
22. M. KHALILIA, S. CHAKRABORTY AND M. POPESCU, "PREDICTING DISEASE RISKS FROM HIGHLY IMBALANCED DATA USING RANDOM FOREST," BMC MEDICAL INFORMATICS AND DECISION MAKING, VOL. 11, (1), PP. 51, 2011.
23. J. V. TU, "ADVANTAGES AND DISADVANTAGES OF USING ARTIFICIAL NEURAL NETWORKS VERSUS LOGISTIC REGRESSION FOR PREDICTING MEDICAL OUTCOMES," J. CLIN. EPIDEMIOL., VOL. 49, (11), PP. 1225-1231, 1996.
24. Z. CHEN, X. ZHANG AND Z. ZHANG, "CLINICAL RISK ASSESSMENT OF PATIENTS WITH CHRONIC KIDNEY DISEASE BY USING CLINICAL DATA AND MULTIVARIATE MODELS," INT. UROL. NEPHROL., VOL. 48, (12), PP. 2069-2075, 2016.
25. S. KARAMIZADEH ET AL, "ADVANTAGE AND DRAWBACK OF SUPPORT VECTOR MACHINE FUNCTIONALITY," IN 2014 INTERNATIONAL CONFERENCE ON COMPUTER, COMMUNICATIONS, AND CONTROL TECHNOLOGY (I4CT), 2014.

DR.RAJESH K S , "PREDICTING GLOBAL ERRORISM ACTIVITIES USING MACHINE LEARNING TECHNIQUES", INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN COMPUTER AND COMMUNICATION ENGINEERING,VOLUME 9, ISSUE 7,JULY2021.