Detection of Chronic Kidney Disease Using Machine Learning Techniques

Tauja K J[1], Sunitha R S[2], Evangeline D[2]

[1]Department of Information Science and Engineering, M.S. Ramaiah Institute of technology, Karnataka, India.
[2]Assistant Professor, Department of Information Science and Engineering, M.S. Ramaiah Institute of technology, Karnataka, India

---

ABSTRACT-*chronic kidney disease it is also called as Chronic Renal Disease, it is a strange working of kidney or a disappointment of renal capacity growing over a time of months or a long time. Constantly, ongoing kidney sickness is recognized during the screening of individuals who are known to be in danger by kidney issues, for example, those with hypertension or diabetes and those with a close family member chronic kidney disease (CKD) patient. Thus, the early expectation is fundamental in battling the infection and to give great treatment. This investigation proposes the utilization of Machine learning techniques like Support Vector Machine (SVM), Naive bayes, Random Forest, Decision Tree classifier. Presently, there are numerous individuals on the planet experiencing chronic kidney infections around the world. Because of the few danger factors like food, climate and expectations for everyday comforts numerous individuals get infections abruptly without comprehension of their condition. Diagnosing of persistent kidney illnesses is by and large intrusive, exorbitant, tedious and frequently hazardous. That is the reason numerous patients arrive at late phases of it without treatment, particularly in those nations where the assets are restricted. Last yield predicts if the individual is having CKD by utilizing least number of highlights. In this project, Naive Bayes, Random Forest, Support Vector Machine and Decision Tree are employed for the disease detection. Comparison of the above algorithms is also done.*

Keywords: CKD, Decision Tree, SVM, Random Forest, Naive Bayes Machine Learning.

I. INTRODUCTION

Because of its high mortality rate, chronic kidney disease (CKD) has received a lot of attention. According to the World Health Organization, chronic kidney disease has become a major concern in developing countries (WHO). CKD is a kidney disease that may be treated in the early stages but leads to renal failure in the later stages. Chronic renal disease claimed the lives of 753 million people worldwide in 2016, including 336 million men and 417 million women. It is classed as a "chronic" illness since the kidney infection develops gradually and lasts a long time, affecting the kidney's function. The amassing of side-effects in the blood prompts the rise of other medical issues, which are related with a few indications, High and low circulatory strain, diabetes, nerve damage, and bone problems are all factors that contribute to cardiovascular disease. Diabetes, pulse, and cardiovascular disease (CVD) are all risk factors for CKD patients. Incidental consequences impair the apprehensive and invulnerable framework in CKD patients, especially of the late stages of the disease. Patients in agricultural countries may come at a late stage, necessitating dialysis or kidney transplants. The glomerular filtration rate (GFR), which depicts kidney function, is used by doctors to diagnose renal disease. GFR is influenced by factors such as age, blood test results, sexual orientation, and other characteristics endured by the patient With respect to GFR esteem, specialists can group CKD into five phases.

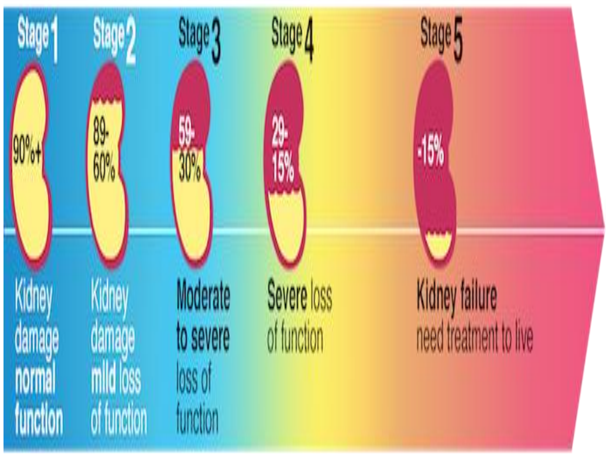


Fig-1: Stages of CKD

II. RELATED WORK

M. P. N. M. Wickramasinghe.et al. [1] provide a method for controlling the condition through the use of a proper food plan, Different techniques, such as Multiclass Decision Trees, neural networks, and Logistic Regression, are used to create classifiers in this research. Depending on the patient's blood potassium level, a potassium zone acceptable is forecasted. A diet spot is suggested by the classification algorithms. To predict Chronic Kidney Disease, H. A. Wibawa.et al [2] designed and evaluated Kernel-based Extreme Learning Machine (ELM). The performance of four portions-based ELMs, namely RBF-ELM, Linear-ELM, Polynomial-ELM, and Wavelet-ELM, is compared to that of regular ELM. The methodologies described above were based on affectability and explicitness measures. Extreme Spiral Basis Function. U. N. Dulhare.et al. [3] used guileless bayes with OneR property selector to extract activity rules based on stages while also predicting CKD, which assists in preventing the progression of chronic kidney infection to later stages. The middle endurance season of past due-stage patients is regarded to be the least difficult, lasting about three years. Exactly assessing the state of the disease. H. Zhang et al. [4] investigated the presentation of Artificial Neural Network (ANN) models in the context of chronic kidney disease (CKD) patients' survival forecasting. Dialysis or kidney transplantation remain the only options for patients with End Stage Renal Disease. Early detection of CKD and genuine treatment can slow or even stop the progression of the disease in the best case scenario. J. Aljaaf et al., [5] reasoned that combining AI computations with predictive analysis yields a clever solution for early infection prediction. Boosting is a group strategy used by data mining models to improve a model's projection. For the most part, AdaBoost and Logit Boost are utilised to examine the display of characterisation classification algorithms.

III. PROPOSED WORK

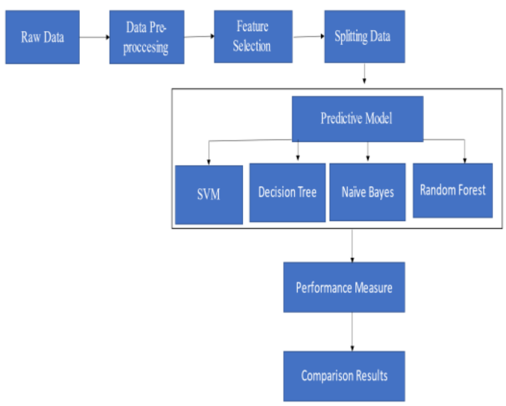


Fig-2: Overall Architecture

Dataset: Machine learning techniques were used to anticipate continuous renal disease in a dataset acquired from the UCI repository. There are 400 patient records included in that dataset and having 24 attributes which are using to build the model. **Data-Prepossessing:** Collect open-source unrefined information on CKD patients from the internet. Because the information we obtained from the web did not include the name of the feature, we initially assigned names to the quality. Missing qualities in the dataset, such as NAs or obvious qualities, are removed using WEKA's "Replace Missing Values" function, which replaces NAs with the trait's mean upsides. Information Reduction: We selected 24 important features in order to build a predictive model. **Training and Testing Dataset:** Datasets for Training and Testing: The dataset is divided into two subsets, each with 24 characteristics. Data for training: The training dataset is derived from the primary dataset and comprises 300 of the 400 entries in the CKD main dataset. Data for testing: 100 out of 400 records from the core CKD dataset were used in the testing.

CLASSIFIERS

Support Vector Machine: "Support Vector Machine" (SVM) is a supervised machine learning algorithm which can be utilized for both classification and regression difficulties. Be that as it may, it is for the most part utilized in arrangement issues. In this algorithm, we plot every data thing as a point in n-dimensional space (where n is number of elements you have) with the worth of each element being the worth of a specific arrange. Then, at that point, we perform order by finding the hyperplane that separate the two classes quite well. Backing Vectors are basically the directions of individual perception. support Vector Machine is a wilderness which best isolates the two classes (hyper-plane/line). The following are the steps for calculating the hyperplane:

1.Set up the training dataset and

2.Set up the SVM border

3. Train the SVM

4. The SVM's characterization of the region

5. Provide a support vector.

Decision Tree: A decision tree is a graphical representation of an explicit choice situation that is used in a predictive model. The root, hubs, and branching decisions are the most important parts of a decision tree. In areas of clinical science when several boundaries are involved with the organisation of informative index, a decision tree is used. Because among all AI calculations, decision tree is the most compressive approach. These are plainly important elements of the informative index. They may also produce the most influential focal point in the general public. The importance of a dataset is obviously connoted by the entropy of a decision tree, and the entropy of a dataset is unmistakably connoted by the entropy of a Decision tree has a disadvantage.

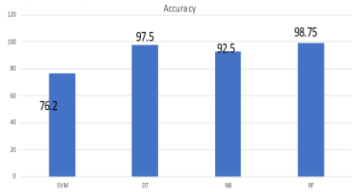
Naïve Bayes: Naive Bayes are probabilistic classifiers, which depend on Bayes Theorem. In Naïve Bayes each worth is stamped autonomous of different qualities and components. Each worth contributes autonomously to the likelihood. The higher the probabilistic worth the higher are the odds of information direct having a place toward that class or classification. Gullible Bayes calculation utilizes the idea of Maximum Likelihood for forecast. This algorithm is quick and can be utilized for making real time predictions such as sentiment analysis.

Random Forest: Random Forest: Random Forest contains hyperparameters that are virtually identical to those of a Decision Tree or a Bagging Classifier. Fortunately, there's no compelling reason to combine a decision tree with a bagging classifier when you can use the random forest's classifier-class instead. You may use the algorithm's regressor to cope with regression tasks using random forest. While developing the trees, the random forest provides more arbitrariness to the model. When splitting a hub, it searches for the best component from an irregular subset of supplies, rather than the primary piece. As a result, there is a large range of options, most of which result in a superior model. In this way, the calculation for parting a hub considers only an irregular subset of the elements in arbitrary timberland. You may make trees even more random by using irregular boundaries for each component rather than searching for the best edges. The Random Forest technique creates numerous decision trees that work together as a classification and regression ensemble. Using a random portion of the training data sets, a number of decision trees are built. The accuracy of findings is improved by using a large number of decision trees. The technique has a relatively short runtime and can handle missing data. The method, not the training data set, is randomized using random forest. The decision class is a type of class that decision trees create.

VI. RESULTS AND CONCLUSIONS

Models were built using a training data set of, which is 70% of the original CKD data set. In terms of parameter correctness, constructed models have been validated using test data that is 30% of the original data. Accuracy was assessed using a confusion matrix in this case. The model with the highest accuracy is the best classification model.

Graphical Representation



CONCLUSION

Chronic Kidney Disease forecast utilizing Machine learning and other classification algorithms. This work is centered around anticipating CKD status of a patient with high precision. Early and Accurate recognition of CKD can accommodate in forestalling further crumbling of patient's wellbeing. In this research we have utilized 24 CKD related characteristics and four arrangement calculations SVM, Decision Tree, Naïve Bayes and Random Forest for foreseeing CKD status of a patient. Same informational collection of 400 individuals was given to both the order calculations and results were acquired. We have looked at the aftereffects of both the calculations based on precision. SVM classifier predicts with an accuracy of 76.2%, though Decision tree Classifier predicts with an accuracy of 97.5%. Naive bayes predicts with an accuracy of 92.5% and Random Forest predicts with an accuracy of 98.75%. Thus, Random Forest gives better accuracy contrast with another Algorithm.

REFERENCES

[1] M. P. N. M. Wickramasinghe, D. M. Perera and K. A. D. C. P. Kahandawaarachchi, "Dietary prediction for patients with chronic kidney disease (CKD) by considering blood potassium level using machine learning algorithms", 2017 IEEE Life Sciences Conference (LSC), Sydney, NSW, 2017, pp. 300-303.

[2] H. A. Wibawa, I. Malik and N. Bahtiar, "Evaluation of Kernel-Based Extreme Learning Machine Performance for Prediction of Chronic Kidney Disease", 2018 2nd International Conference on Informatics and Computational Sciences (ICICoS), Semarang, Indonesia, 2018, pp. 1-4.

[3] U. N. Dulhare and M. Ayesha, "Extraction of action rules for chronic kidney disease using Naïve bayes classifier",2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Chennai, 2016, pp. 1-5.

[4] H. Zhang, C. Hung, W. C. Chu, P. Chiu and C. Y. Tang, "Chronic Kidney Disease Survival Prediction with Artificial Neural Networks", 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Madrid, Spain, 2018, pp. 1351-1356.

[5] J. Aljaaf et al., "Early Prediction of Chronic Kidney Disease Using Machine Learning Supported by Predictive Analytics", 2018 IEEE Congress on Evolutionary Computation (CEC), Rio de Janeiro, 2018, pp. 1-9.