IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Comparing Machine Learning Approaches For Chronic Kidney Disease Prediction

[1]E. Tejasree, [2]Dr. Kondapalli Venkata Ramana
[1]M. Tech, [2]Associate Professor
[1,2]Dept of CS&SE, AUCE, AU, Visakhapatnam

ABSTRACT

The field of biosciences have progressive to a higher extent and have generated large amounts of information from Electronic Health Records. This have given rise to the acute need of knowledge generation from this enormous amount of data. Data mining methods and machine learning play a major role in this aspect of biosciences. Chronic Kidney Disease (CKD) is a condition in which the kidneys are damaged and cannot filter blood as they always do. A family history of kidney diseases or failure, high blood pressure, type 2 diabetes may lead to CKD. This is a lasting damage to the kidney and chances of getting worse by time is high. The very common problems that results due to a kidney failure are heart diseases, anemia, bone diseases, high potassium, and calcium. The worst-case situation leads to complete kidney failure and necessitates kidney transplant to live. An early detection of CKD can increase the quality of life to a greater extent. This calls for good prediction algorithm to predict CKD at an earlier stage. Literature shows a wide range of machine learning algorithms employed for the prediction of CKD. This paper uses data preprocessing, data transformation and various classifiers to predict CKD and proposes best Prediction framework for CKD. The results of the framework show promising results of better prediction at an early stage of CKD.

Keywords: Chronic Kidney Disease, Machine Learning Approaches, Kidney Disease Prediction.

1. INTRODUCTION

The disability of the kidneys to perform their regular blood filtering function and others is called Chronic Kidney Disease (CKD). The term “chronic” describes the slow degradation of the kidney cells over a long period of time. This disease is a major kidney failure where the kidney sans blood filtering process and there is a heavy fluid buildup in the body. This leads to alarming increase of potassium and calcium salts in the body. Existence of high levels of these salts result in various other ailments in the body. The prime job of kidneys is to filter extra water and wastes from blood. The efficient functioning of this process is important to balance the salts and minerals present in our body. The rich balance of salts is necessary to control blood pressure, activate hormones, build red blood cells, etc. A high concentration of calcium leads to various bone diseases and cystic ovaries in women. CKD also may lead to sudden illness or allergy to certain medicines. This state is called as Acute Kidney Injury (AKI). An increased blood pressure may lead to heart problems and heart attacks. CKD in many cases leads to permanent dialysis or kidney transplants.

1.1 Background & Motivation

A history of kidney disease in the family also leads to high probability of CKD. Literature shows that almost one out of three people diagnosed with diabetes have CKD. Literature also presents evidences of early identification and care of CKD can improve the quality of the patient’s life. Prediction algorithms in machine learning can be intelligently used to predict the occurrence of CKD and presents a method of early medication. The detailed review on literature shows the application of various machine learning algorithms to predict CKD. This paper tries to predict CKD using the classifiers like Decision Tree, Random Forest and Support Vector Machine and suggested best prediction model.

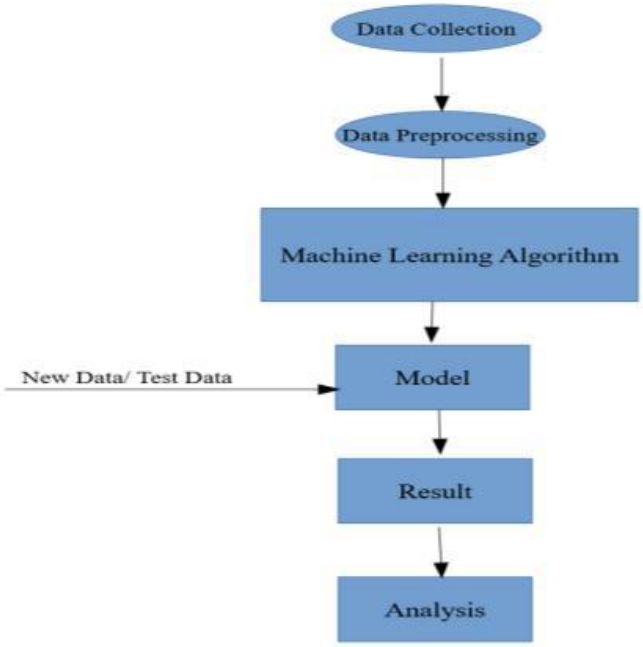

Figure 1: Stages of the Knowledge Data Discovery process

The disease is called "chronic" because the damage to your kidneys happens slowly over a long period of time. The main causes are damaged blood vessels of the kidneys due to High Blood Pressure and Diabetes. The CKD is also called a chronic kidney failure where according current medical statistics the 10% of the population worldwide is affected by CKD. There were approximately 58 million deaths in the year of 2019 worldwide. According to the World Health Organization (WHO), 35 million attributed to chronic diseases. Currently it is estimated that one in five men, and one in four women aged 65 through 74 are going to be affected by CKD worldwide. Diagnosing CKD usually starts with clinical data, lab tests, imaging studies and finally biopsy. In this study, by using the machine learning techniques, we are proposing cheap, simple and non-invasive tests that can be performed easily.

2. LITERATURE REVIEW

The healthcare domain is one of the prominent research fields in the current scenario with the rapid improvement of technology and data. It is difficult to handle the huge amount of data of the patients. There are a lot of procedures for the treatment of multiple diseases across the world. Machine Learning is an emerging approach that helps in prediction, diagnosis of a disease.

M. P. N. M. Wickramasinghe.et al(1) presents a methodology to control the disease using a suitable diet plan. In this research classifiers are constructed using different algorithms like Multiclass Decision Jungle, Multiclass Decision Forest, Multiclass Neural Network and Multiclass Logistic Regression. An allowable potassium zone is predicted depending on the blood potassium levels of the patient. The classification algorithms recommend a diet place based on the predicted potassium zone. *H. A. Wibawa.et al*(2) proposed and evaluated Kernel-based Extreme Learning Machine(ELM) to predict Chronic Kidney Disease. Performance of four kernels-based ELM, namely RBF-ELM, Linear-ELM, Polynomial-ELM, Wavelet-ELM are compared with the performance of standard ELM. The above methodologies were compared on metrics of sensitivity and specificity. Radial Basis Function – Extreme Learning Machine (RBF-ELM) showed higher prediction rates.

CKD increases the risk factors of cardiovascular disease (CVD) like hypertension, diabetes mellitus, dyslipidemia, and metabolic syndrome. CKD also leads to End Stage Renal Disease (ESRD) which has no cure. *U. N. Dulhare.et al*(3) extracted action rules based on stages but also predicted CKD by using naïve bayes with One R attribute selector which helps to prevent the advancing of chronic renal disease to further stages.

It is said that the median survival time of past due-stage patients is simplest approximately three years. Evaluating exactly the condition of sufferers is of incredible importance as it might substantially assist to decide appropriate care, medications or medical interventions wished, which amongst them have a complicated interrelationship and have an impact on the results of the person patient. *H. Zhang.et al*(4) investigated the performance of Artificial Neural Network (ANN) models while applying to the survivability prediction on Chronic Kidney Disease (CKD) patients.

Dialysis or Kidney transplant stays the only option for a patient with End Stage Renal Disease (ESRD). The progression of the disease can be slowed down or even stopped in a favorable case
by early prediction of CKD and proper treatments with diet. *J. Aljaaf et al.*, (5) concluded that application of machine learning algorithms with predictive analytics proves to be an intelligent solution for early prediction of the disease.
Data mining models project ensemble techniques called Boosting which enhances the prediction of a model. AdaBoost and LogitBoost are generally used to compare the performance of classification algorithms. *Arif-Ul-Islam. et al*(6) analyzed the performance of boosting algorithms for detecting CKD and derived rules illustrating relationship among the various attributes of CKD. The paper used Ant-Miner machine learning algorithm along with Decision tree to derive rules.
Datamining methods are used to generate decisions by elicitating hidden information from chronic disease datasets. This calls of storage and manipulation of large amounts of structured, unstructured, and semi structured data. The role of big data in for the same is very important. ***G.*** *Kaur.et al*(7) predicted chronic kidney disease using various data mining algorithms in Hadoop environment. Classifiers like KNN (K-Nearest Neighbor) and SVM (Support Vector Machine) are used in the research.
Levels of creatinine, sodium, urea in blood play an important role in deciding the survival prediction or the need for kidney transplantation in patients undergoing dialysis and becoming worser.*V.Ravindra.et al*(8) used simple K-means algorithm to elicit knowledge about the interaction between many of these CKD parameters and patient survival. He concluded that the clustering procedure predicts the survival period of the patients who undergo the dialysis procedure. For CKD prediction, *R.Devika et al*(9) examined the performance of Naive Bayes, K-Nearest Neighbour (KNN) and Random Forest classifiers based on accuracy, preciseness and execution time for. *P.Panwong.et al*(10) created a classification model for predicting transitional interval of Kidney disease stages 3 to 5 and also used Decision tree, K-nearest neighbor, Naïve Bayes and Artificial neural networks for eliciting the knowledge and creating classification model with the selected set of attributes. *S.Vijayarani.at.al* (11) predicted kidney diseases by using Support Vector Machine (SVM) and Artificial Neural Network (ANN). The research compared the performance of the above two algorithms on accuracy and execution time. *Misir R, et al*(12) used feature selection algorithms to identify set of features that efficiently predict kidney diseases. The reduced feature set results in reduced costs, saves time and reduced uncertainty.

Kidney damage due to diabetes is chronic and a slow process, but has significant effects on the patient. High glucose levels in blood disturbs the kidney from functioning effectively. *Bharathiet.al*(16) has applied association rule mining to predict diabetes mellitus in a given dataset by generating summarization rules.

3. PROBLEM DEFINITION

Nowadays, health care industries are providing several benefits like fraud detection in health insurance, availability of medical facilities to patients at inexpensive prices, identification of smarter treatment methodologies, and construction of effective healthcare policies, effective hospital resource management, better customer relation, improved patient care and hospital infection control. Disease detection is also one of the significant areas of research in medical. There is no automation for chronic kidney disease prediction.

Description of Dataset

The clinical data of 400 records considered for analysis has been taken from UCI Machine Learning Repository. The data obtained after cleaning and removing missing values is 220. There are 25 attributes in the dataset. The numerical attributes include age, blood pressure, blood glucose random, blood urea, serum creatinine, sodium, potassium, hemoglobin, packaged cell volume, WBC count, RBC count. The nominal attributes include specific gravity, albumin, and sugar. It also includes RBC, pus cell and pus cell clumps, bacteria, hypertension, diabetes mellitus, coronary artery disease, appetite, pedal edema, anemia, and class. Number of Instances: 400 Number of Attributes: 25 Class: {CKD, NOTCKD} Missing Attribute Values: yes, Class Distribution: [63% for CKD] [37% for NOTCKD]
The most popular R Programming Data analytics tool has been used to construct the prediction framework.

	age	bp	sg	al	su	rbc	pc	pcc	ba	bgr	bu	sc	sod
1	48	80	1.020	1	0	?	normal	notpresent	notpresent	121	36	1.2	?
2	7	50	1.020	4	0	?	normal	notpresent	notpresent	?	18	0.8	?
3	62	80	1.010	2	3	normal	normal	notpresent	notpresent	423	53	1.8	?
4	48	70	1.005	4	0	normal	abnormal	present	notpresent	117	56	3.8	111
5	51	80	1.010	2	0	normal	normal	notpresent	notpresent	106	26	1.4	?
6	60	90	1.015	3	0	?	?	notpresent	notpresent	74	25	1.1	142

Figure 2: Chronic Kidney Disease Dataset

4. IMPLEMENTATION:

Our Aim is to predict the chronic kidney disease using the machine learning algorithms. Chronic kidney disease (CKD) means your kidneys are damaged and cannot filter blood the way they should. The disease is called "chronic" because the damage to your kidneys happens slowly over a long period of time. This damage can cause wastes to build up in your body. CKD can also cause other health problems.10% of the population worldwide is affected by chronic kidney disease (CKD), and millions die each year because the doctors are unable diagnose the disease.

The system is automation for predicting the CKD. The system is a Real-world web-based application that can be used by many hospitals. It assumes variables are independent of each other. The algorithms were easy to build and works well with huge data sets. It has been used because it makes use of small training data to estimate the parameters important for classification. It performs well in multiple class prediction. When assumption of independence holds, a various classifiers perform better compare to other models like logistic regression and you need less training data.

Algorithm:

Input: Chronic Kidney Disease Dataset

Output: High Accuracy prediction Framework

Step1: Input data

Step2: Pre-process the data

Step 2.1: Convert Categorical values to numerical values

Step 2.2: Replace numerical missing values by Mean

Step2.3: Replace Categorical missing values by Mode

Step3: Construct Classifier Models

Step3.1: Construct Decision Tree Model

Step3.2: Construct Random Forest Model

Step3.3: Construct SVM model

Step3.4: Construct KNN Model

Step3.5: Construct Voting Classifier Model.

Step3.6: Construct Linear Regression Model.

Step 4: Check the accuracy of the constructed models using confusion matrix

Step 5: Decide the best prediction model for CKD.

4.1 Decision Trees: Decision Trees are supervised learning algorithms that can be used for classification and regression tasks. They form tree-like structures by analyzing the data and learning simple decision rules. Unlike parametric algorithms like Logistic Regression, decision trees do not learn parameters, but rather learn decision rules that are arranged in a hierarchical manner to make predictions. One advantage of decision trees is their interpretability. They provide intuitive insights into the decision-making process by showing which features contribute to the classification/regression outcome. This makes decision trees particularly useful when it's important to understand the factors influencing the target variable.

In our case, we use a classification decision tree, where decision rules are present in the nodes of the tree. Based on the outcome of a decision node, a specific branch is chosen, leading to another node with a different decision to make. This recursive process continues until reaching the leaves of the tree, which contain class labels attributed to the data items.

Decision trees are well-suited for problems with multiple target classes, as is the case in our scenario. The efficiency of decision trees for large datasets is proportional to the logarithm of the number of data points in the training set. Additionally, decision trees are considered "white-box" models, as their structure allows for easy visualization and understanding of the classification process.

However, decision trees can become complex with large datasets, making them prone to overfitting the training set. Learning a decision tree in polynomial time efficiently is challenging, and there is no known efficient algorithm for this task. Moreover, decision trees tend to perform poorly when dealing with imbalanced datasets, unless proper class balancing methods are applied beforehand.

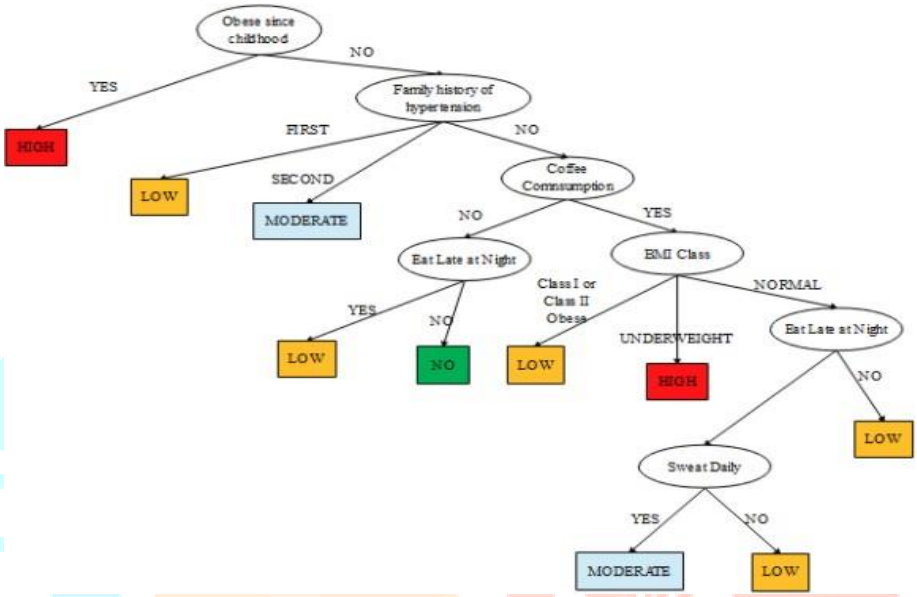

Figure 3: Decision Trees

4. 2 Random Forests

Random Forest is a popular ensemble approach that aims to reduce the correlation between the predictions of decision trees in order to improve the model's accuracy. Instead of each tree making predictions based on all the features, Random Forest randomly selects a subset of features to use at each node. By doing so, the trees are decorrelated from one another, reducing the chances of making the same type of errors.

In Random Forest, multiple decision trees are trained independently, each on a random subset of the training data and a random subset of the features. This helps to prevent overfitting and improves the model's generalization ability. In addition, Random Forest uses a technique called feature bagging, which helps to reduce the impact of strongly relevant features on the overall predictions.

By using Random Forest, we can improve the accuracy and robustness of our model, making it less likely to make the same type of errors. This approach is particularly useful for complex datasets with many features and where overfitting is a concern. However, it should be noted that Random Forests are still prone to overfitting if the number of trees is too large or if the depth of the trees is too high. In addition, Random Forests can be computationally expensive, especially when dealing with large datasets.

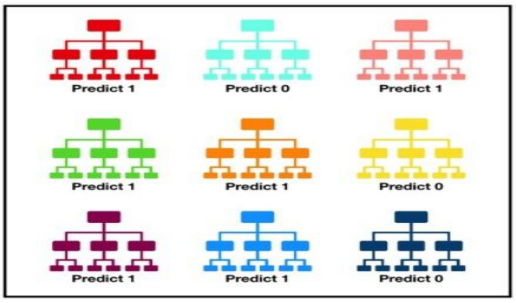

Figure 4: Random Forest

4.3 Support Vector Machines (SVM)

Support Vector Machines (SVMs) are a type of supervised machine learning algorithm that can be used for classification and regression tasks, and are suitable for both linear and non-linear data. The goal of SVMs is to find an optimal hyperplane in N-dimensional space, which separates the data points and maximizes the margin between the closest support vectors. Support vectors are the data points closest to the hyperplane and are used for training, while the margin is the distance between the two parallel lines passing through the support vectors on either side of the hyperplane.

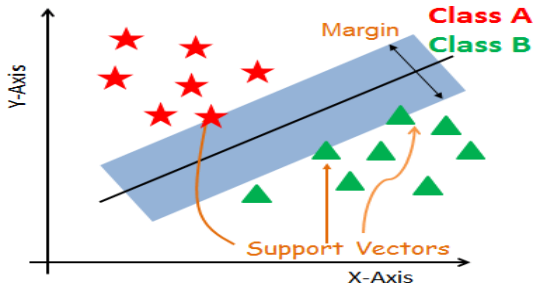

Figure 5: Support Vector Machine

For non-linearly separable data, SVMs use the Kernel Trick to transform the data into a higher dimensional space for classification. SVMs are efficient for classifying high dimensional data, and they are memory-efficient because only the support vectors are used for training. However, SVMs perform poorly when dealing with noisy data or datasets with overlapping classes, and they are not very efficient for large datasets because the training time increases as the dataset size increases.

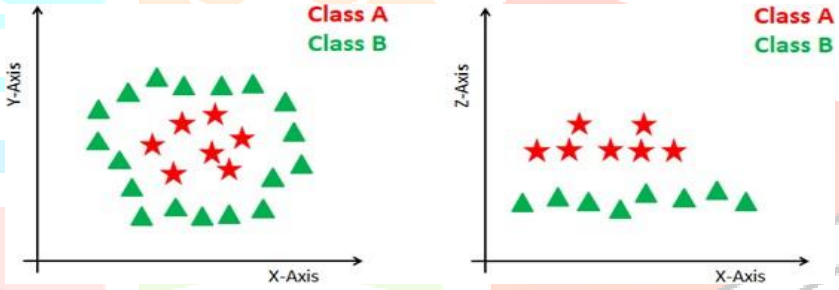

Figure 6: SVM on non-linearly separable data

4.4 K-Nearest Neighbour

K - Nearest Neighbor (KNN) is a very simple, easy to understand, versatile and one of the topmost machine learning algorithms. KNN used in the variety of applications such as finance, healthcare, political science, handwriting detection, image recognition and video recognition. In Credit ratings, financial institutes will predict the credit rating of customers. In loan disbursement, banking institutes will predict whether the loan is safe or risky. In political science, classifying potential voters in two classes will vote or will not vote. kNN algorithm used for both classification and regression problems. kNN algorithm based on feature similarity approach.

KNN Algorithm

In KNN, K is the number of nearest neighbors. The number of neighbors is the core deciding factor. K is generally an odd number if the number of classes is 2. When K=1, then the algorithm is known as the nearest neighbor algorithm. This is the simplest case. Suppose P1 is the point, for which label needs to predict. First, you find the one closest point to P1 and then the label of the nearest point assigned to P1.

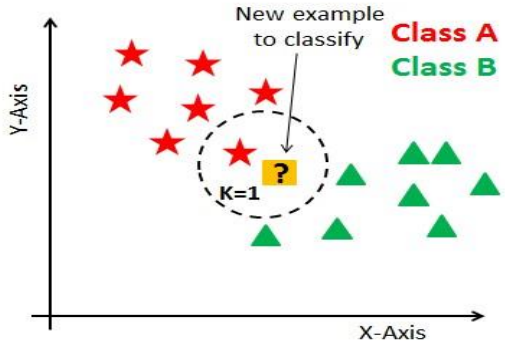

Figure 7: Example for KNN

Suppose P1 is the point, for which label needs to predict. First, you find the k closest point to P1 and then classify points by majority vote of its k neighbors. Each object votes for their class and the class with the most votes is taken as the prediction. For finding closest similar points, you find the distance between points using distance measures such as Euclidean distance, Hamming distance, Manhattan distance and Minkowski distance.

4.5 Logistic Regression

Logistic regression is another technique borrowed by machine learning from the field of statistics.

Logistic Function

Logistic regression is named for the function used at the core of the method, the logistic function.

The logistic function, also called the sigmoid function was developed by statisticians to describe properties of population growth in ecology, rising quickly and maxing out at the carrying capacity of the environment. It is an S-shaped curve that can take any real-valued number and map it into a value between 0 and 1, but never exactly at those limits.

$1 / (1 + e^{-value})$

Where 'e' is the base of the natural logarithms (Euler's number or the EXP() function in your spreadsheet) and value is the actual numerical value that you want to transform. Below is a plot of the numbers between -5 and 5 transformed into the range 0 and 1 using the logistic function.

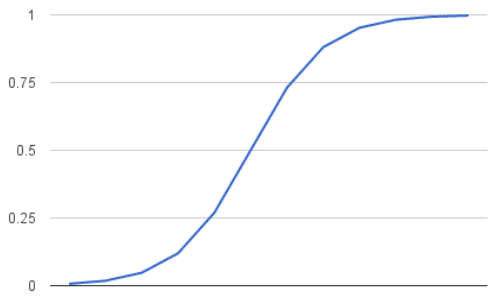

Figure 8: Logistic Regression

4.6 Voting Classifier

A Voting Classifier is a machine learning model that trains on an ensemble of numerous models and predicts an output (class) based on their highest probability of chosen class as the output. It simply aggregates the findings of each classifier passed into Voting Classifier and predicts the output class based on most of the voting. The idea is instead of creating separate dedicated models and finding the accuracy for each them, we create a single model which trains by these models and predicts output based on their combined majority of voting for each output class.

Voting Classifier supports two types of voting's.

1. ***Hard Voting:*** *In hard voting, the predicted output class is a class with the highest majority of votesi.e the class which had the highest probability of being predicted by each of the classifiers. Suppose three classifiers predicted the* output class (A, A, B)*, so here the majority predicted* A *as output. Hence* A *will be the final prediction.*

2. ***Soft Voting:*** *In soft voting, the output class is the prediction based on the average of probability given to that class. Suppose given some input to three models, the prediction probability for class* A = (0.30, 0.47, 0.53) *and* B = (0.20, 0.32, 0.40)*. So, the average for class* A is 0.4333 *and* B is 0.3067*, the winner is clearly class* A *because it had the highest probability averaged by each classifier.*

5. RESULTS:

The prediction results have been evaluated using following parameters:

Precision: It is the fraction of retrieved data that are useful for the query.
Recall: It is the fraction of data that are relevant for the query which is effectively retrieved.
F-measure: It is measure that sums up precision and recall.

In this we perform prediction for predicting the dataset which are related to the new document for the future usage. This predicted dataset is used to build a model. That predictive model is then used on current dataset to predict what will happen next or to suggest actions to take for optimal outcomes.

Accuracy Calculations

Using classifications algorithms to fit the predicted data, this would run our data over the current data and gives us the accuracy of the classification. The classification.

Accuracy is the ratio of number of correct predictions to the total number of input samples.

Accuracy = Number of correct predictions / Total number of predictions made

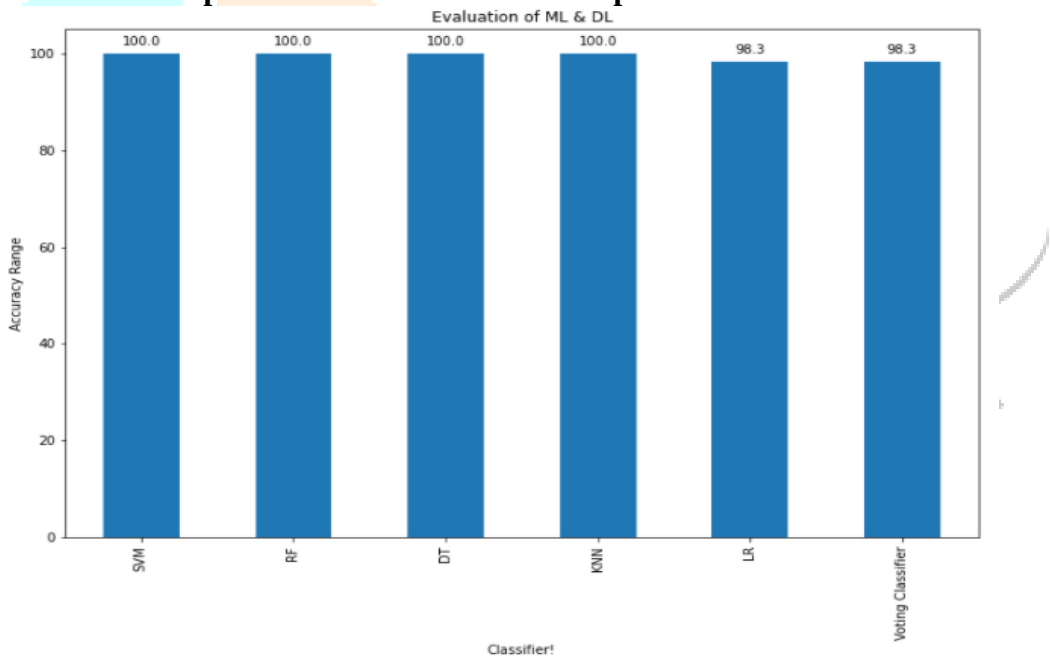

Figure 9: Accuracy Comparison

6. CONCLUSION

This project presented a prediction algorithm to predict CKD at an early stage. The dataset shows input parameters collected from the CKD patients and the models are trained and validated for the given input parameters. Decision tree, Random Forest, KNN, Linear Regression and Support Vector Machine, Voting Classifier learning models are constructed to carry out the diagnosis of CKD. The performance of the models is evaluated based on the accuracy of prediction. The results of the research showed that SVM, Decision Tree, KNN, Random Forest Classifier model better predicts CKD in comparison to other algorithms. The comparison can also be done based on the time of execution, feature set selection as the improvisation of this research. The data distribution has properly covered the whole domain in CKD, but the general attributes like appetite, anaemia and pedal oedema are biased towards CKD. It is easy to achieve an accurate prediction using this data set but in the general context, it may lead to false positives as observed in the recall column. Further, the missing values which were completely missed at random made it impossible to achieve a perfect accuracy without filling them from a collaborative imputer instead of a constant.

Considering the medical importance of the attributes, some of them have a lesser co-relation compared to others because of the stage they appear in the patient. When training the models, it makes a huge impact on the accuracy. After training the model, it clearly shows that tree structures have higher accuracy than other classification algorithms, which can be justified from the distribution of the data set since the selected attributes have a clearer separation in the class except for serum creatinine attribute. Finally, when selecting the algorithm, some trained models have a bias towards some. Considering the causes of change of the nominal values of them, it has many different possibilities apart from CKD. Therefore, it motivates to rely less on one attribute and consider more when making the decision and based on that the extra tree classifier has been selected.

Future Scope

We aim to validate our results by using big dataset or compare the results using another dataset that contains the same features. Also, in order to help in reducing the prevalence of CKD, we plan to predict if a person with CKD risk factors such as diabetes, hypertension, and family history of kidney failure will have CKD in the future or not by using appropriate dataset. In the future course of this study one can try to further improve the two-class classification accuracy by evaluating some hybrid or ensemble techniques, in addition to this a subset of features can be extracted from the complete medical data-set of chronic kidney disease of twenty-four parameters (features) without effecting the performance of the classification process, so that the financial burden a patient must bear for undergoing various clinical tests can be reduced.

7. REFERENCE:

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. Aljaafet 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.
6. Arif-Ul-Islam and S. H. Ripon, "Rule Induction and Prediction of Chronic Kidney Disease Using Boosting Classifiers, Ant-Miner and J48 Decision Tree," 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), Cox'sBazar, Bangladesh, 2019, pp. 1-6.
7. G. Kaur and A. Sharma, "Predict chronic kidney disease using data mining algorithms in hadoop," 2017 International Conference on Inventive Computing and Informatics (ICICI), Coimbatore, 2017, pp. 973-979.
8. N. Tazin, S. A. Sabab and M. T. Chowdhury, "Diagnosis of Chronic Kidney Disease using effective classification and feature selection technique," 2016 International Conference on Medical Engineering, Health Informatics and Technology (MediTec), Dhaka, 2016, pp. 1-6.
9. V. Ravindra, N. Sriraam and M. Geetha, "Discovery of significant parameters in kidney dialysis data sets by K-means algorithm," International Conference on Circuits, Communication, Control and Computing, Bangalore,2014,pp.452-454.
10. R. Devika, S. V. Avilala and V. Subramaniyaswamy, "Comparative Study of Classifier for Chronic Kidney Disease prediction using Naive Bayes, KNN and Random Forest," 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2019, pp. 679-684.
11. P. Panwong and N. Iam-On, "Predicting transitional interval of kidney disease stages 3 to 5 using data mining method," 2016 Second Asian Conference on Defence Technology (ACDT), Chiang Mai, 2016, pp. 145-150.
12. S. Vijayarani, S. Dhayanand, "KIDNEY DISEASE PREDICTION USING SVM AND ANN ALGORITHMS", International Journal of Computing and Business Research (IJCBR), vol. 6, no. 2, 2015.
13. Misir R, Mitra M, Samanta RK. A Reduced Set of Features for Chronic Kidney Disease Prediction. J Pathol Inform.2017.

14. "UCI Machine Learning Repository: Chronic_Kidney_DiseaseDataSet", Archive.ics.uci.edu, 2015. Available:http://archive.ics.uci.edu/ml/datasets/Chronic_Kidney_Disease
15. B.Bharathi, S.Prince Mary (2019), Neural Computation based general disease prediction model, International journal of Recent Technology and Engineering, vol8(2),pp 5646-5649.
16. Kamalesh M.D., Predictind the risk of diabetes mellitus to subpopulations using association rule mining, Proceedings of the International Conference of SoftComputing systems, Advances in Intelligent Systems and Computing col.397, Springer (2016)
17. Revathy,B.Parvathavarthini,Shiny Caroline, Decision Theory, an Unprecedented Validation Scheme for Rough-Fuzzy Clustering, International Journal on Artificial Intelligence Tools, World Scientific Publishing Company ,Vol.25,No.2,2016.