2020 IEEE International Conference for Innovation in Technology (INOCON)
Bengaluru, India. Nov 6-8, 2020

Implementation of Machine Learning Algorithms to Detect the Prognosis Rate of Kidney Disease

F. M. Javed Mehedi Shamrat
Department of Software Engineering
Daffodil International University
Dhaka, Bangladesh
javedmehedicom@gmail.com

Pronab Ghosh
Department of Computer Science & Engineering
Daffodil International University
Dhaka, Bangladesh
pronab1712@gmail.com

Mahbubul Hasan Sadek
Department of Computer Science & Engineering
Daffodil International University
Dhaka, Bangladesh
hasan15-4966@diu.edu.bd

Md. Aslam Kazi
Department of Computer Science & Engineering
Daffodil International University
Dhaka, Bangladesh
kaziaslam5490@gmail.com

Shahana Shultana
Department of Computer Science & Engineering
Daffodil International University
Dhaka, Bangladesh
shahanashomi246@gmail.com

Abstract— **The chronic kidney disease is the loss of kidney function. Often time, the symptoms of the disease is not noticeable and a significant amount of lives are lost annually due to the disease. Using machine learning algorithm for medical studies, the disease can be predicted with a high accuracy rate and a very short time. Using four of the supervised classification learning algorithms, i.e., logistic regression, Decision tree, Random Forest and KNN algorithms, the prediction of the disease can be done. In the paper, the performance of the predictions of the algorithms are analyzed using a pre-processed dataset. The performance analysis is done base on the accuracy of the results, prediction time, ROC and AUC Curve and error rate. The comparison of the algorithms will suggest which algorithm is best fit for predicting the chronic kidney disease.**

Keywords— Logistic Regression, Decision Tree, Random Forest, K-Nearest Neighbors, Accuracy

I. INTRODUCTION

Chronic kidney disease is considered as one of the most deadly diseases globally. It is a progressive harmful disease, is propagated in different situations. With the right tools, however, it is possible to make the correct decision in order to do the proper treatment. In this regard, machine learning classification algorithm plays a vital part [1]. Using various machine learning algorithms, the prognosis rate of the disease can to studied, evaluated and proper actions can the taken based on the results and discussions. Machine learning classification algorithms categorizes data to create different classes. In supervised learning [2], Classification is one of the most important characteristics. There are several supervised learning classification algorithm, but in this study, Logistic Regression, Decision Tree, Random Forest, K-Nearest Neighbors algorithms are implemented. These algorithms, each have unique properties of their own that gives them their own recognitions[3]. Each algorithm has different accuracy or execution rate. In this study, the mentioned machine learning algorithms will be implemented on the data of chronic kidney patients in order to determine which algorithms gives the best result to predict the disease.

II. LITERATURE REVIEW

Machine learning algorithms have become a very essential tool in not just the field of math and engineering but medicine as well. Using machine learning algorithms, predictions are done at a regular basis in order to take quick and necessary actions. Mostly, to improve performance and accuracy, algorithms are implemented.

In [4], authors studied the human activities. Using the data acquired from UCI machine learning repository, human activity vs. the smart-phone based human activity is studied by implementing the logistic regression, KNN, and random forest regress among various others. The accuracy rates are compared to show that logistic regression gives the most accurate prediction.

The authors examined the result of sonographic and clinical variables to predict the acute appendicitis in minors in [5]. In this paper, the decision tree algorithm is implemented to identify discriminatory variables and the receiver operative characteristics (ROC) curve is analyzed to calculate performance.

Based on the dengue epidemic in Singapore in 2013, the authors implanted random forest algorithm on the data available on dengue, population entomology and environment, to predict the risk of transmission of the virus in 1 square kilometer grid in the paper [6].

In the paper [7] the image processing data of brain tumor is utilized to actualize KNN algorithm. Here the algorithm gives the expectation of the presence of cerebrum tumor in the MRI image of the brain. The result is analyzed to determine the accuracy of the algorithm for detection.

For the research gap, the Logistic Regression, Decision Tree, Random Forest, K-Nearest Neighbors algorithms are implemented on the data of chronic kidney disease patients to predict the disease. Here the algorithms are implemented on preprocessed data. The result comparison of all the algorithms are done successfully based on the same dataset. The performance comparison is done base on criteria such as accuracy of the detection rate, execution time, ROC and AUC curves, and error rate to determine the most efficient algorithm for chronic kidney disease detection.

978-1-7281-9744-9/20/$31.00 ©2020 IEEE

III. Research Methodology

A. System Overview:

This methodology works with the theoretical capacity of research-based work. It will provide a piece of clear information about the concept of work. To make it simpler, research subject and instrumentation will shortly be identified very first. In data mining, data is considered to be the heart. The algorithms are them implemented on the data after the data are being preprocessed.

The data used for the study is collect from an online data repository. The data is preprocessed to make it suitable for the implementation of any machine learning classification algorithms. The dataset is processed further to handle missing data. Feature scale and classification is done before implementation of algorithms.

The Logistic Regression algorithm, Decision Tree algorithm, Random Forest algorithm, K-Nearest Neighbors algorithm are implemented in the pre-processed data. From the outcome of each algorithm, performance evaluation is done. The detecting of the best-fit algorithm accurately is used to recommend the best tool.

The following Fig. 1 illustrates the methodology employed to perform a comparative analysis to recommend the best algorithm to detect the kidney disease.

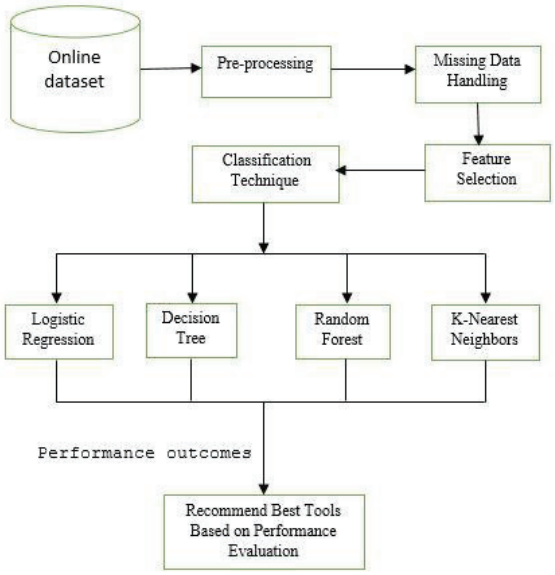

Fig. 1. Functional Blocks of kidney disease prediction System

B. Evaluation Criteria on Performance Measure Indices:

Some performance indicators such as accuracy, recall, F1-score, accuracy, specification (space), false-positive rate, false-negative rate, and negative predictive values are used to represent the four different learning strategy models. The results of the performance measurement indicators depend on TP, TN, FP and FN. [13]

True Positive (TP) = The number of examples exactly identified with Chronic Kidney Disease.

False Positive(FP) = The number of instances incorrectly classified with Chronic Kidney Disease.

True Negative(TN) = The number of instances exactly identified with Chronic Kidney Disease.

False Negative(FN) = The number of instances exactly identified with Chronic Kidney Disease.

$$\text{Accuracy (Acc)} = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (1)$$

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

$$\text{Recall} = \frac{(TP)}{(TP+FN)} \quad (3)$$

$$\text{F1-score} = \frac{2(\text{Precision X Recall})}{(\text{Precision}+\text{Recall})} \quad (4)$$

$$\text{Sensitivity (Sen)} = \frac{(TP)}{(TP+FN)} \text{ [14]} \quad (5)$$

$$\text{Specificity (Spec)} = \frac{(TN)}{(TN+FP)} \quad (6)$$

$$\text{false positive rate} = \frac{FP}{FP+TN} \quad (7)$$

$$\text{False Negative Rate} = \frac{100 \text{ x } FN}{(TP+FN)} \quad (8)$$

$$\text{Negative predictive value} = \frac{TN}{(TN+FN)} \quad (9)$$

IV. Implementation

A. Tool and Language:

In this study, we used Google Colab as a model building tool and python 3.7, a programming language, has been used to work over the UCI dataset.

B. Dataset Collection:

In specific cases, data is deliberated as the first and foremost part of the research. Since the technique of machine learning is thought the center of the data to obtain the expected outcome, we tried to collect it locally but due to certain constraints, we did not get it. As a result, a considerable amount of patient records was collected from databases to obtain unique results. In this study, the chronic kidney disease dataset was applied from the UCI Machine Learning Repository [8] to predict deadly disease. There are two groups of illnesses shown in Fig. 2.

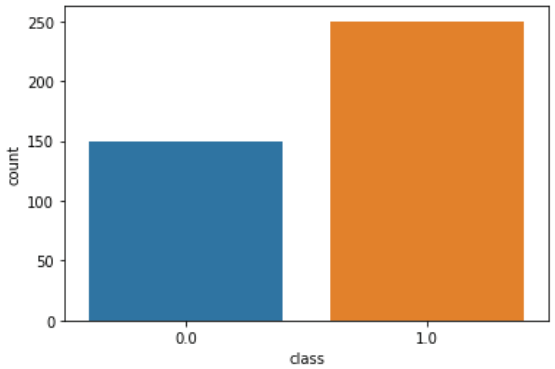

Fig. 2. The Outcomes of Disease Attributes

In this study, 400 data has been collected from South India with 25 individual features [9] such as age, blood pressure, cholesterol, urea in the blood, appetite, red blood cells, sodium, etc. The age range of patients is within the range of 2-90 years. We have some issues in this dataset, including missing interest. There are two classes: Kidney Disease, KD (250) and non-KD (150). One of the most popular label encoding techniques [10] has been founded to

convert all categorical features into numbers. For example, "non-KD" and "KD" represent 0 and 1.

C. Data Pre-processing:

A chart or graph is used to indicate a large amount of complex data which is easier than trimming a spreadsheet. After collecting several raw data from one of the most popular repositories, the dataset had to be applied to the preprocessing section. In this region, many missing values are identified to meet data demand. Almost every column has a drastic number of missing values that are found in Fig. 3.

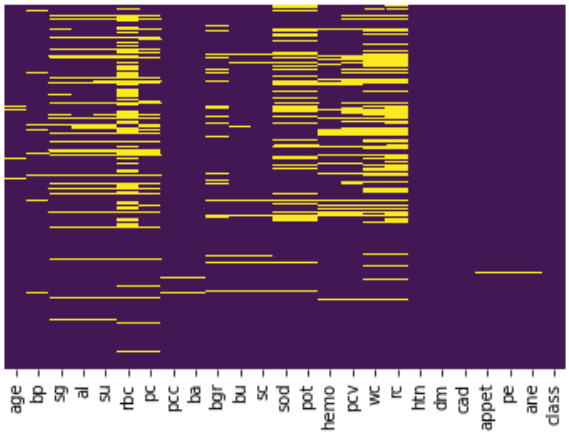

Fig. 3. Missing Values of the Exhibited Dataset

We collect a lot of valid data from various repositories to get an accurate forecast rate. While the dataset contains many missing values, it creates complex situations for us to identify KD patients. This is considered to be one of the biggest challenges among them. There are two recognized methods such as substitution by mean and substitution by median to fill in the missing values. In this System, we have used one of the most popular techniques 'replacement by median'. After using this recognized technique, no missing values were found in the nearest figs. 4.

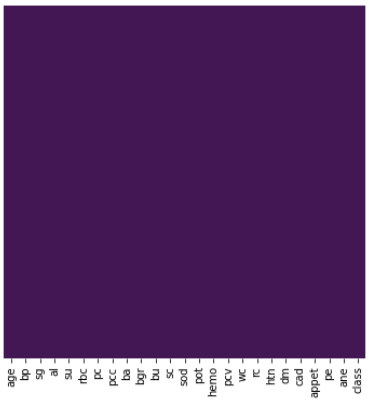

Fig. 4. No Missing Values in Chronic kidney Disease Dataset

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. [11] [12]. 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.

D. Logistic Regression:

Logistic regression is a model [15] called statistical machine learning technique that helps to classify classified variables [16] and used for binary classification. Assuming the probability of a particular class, LR creates a model that distinguishes between samples [17-18]. One or more variables are used to determine the expected outcome. If p is the probability then, the logit function for p is defined as [19]:

$$Logit(p) = ln(\frac{p}{1-p}) \quad (10)$$

Figure 5 shows the implementations of the logistic regression algorithm.

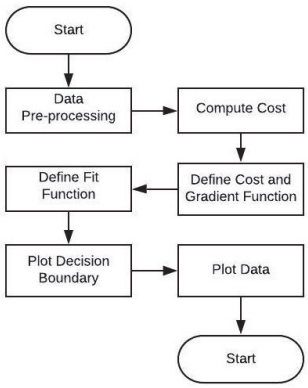

Fig. 5. Implementation of the logistic regression

E. Decision Tree:

The technique of decision tree is deliberated as one of the most powerful tools [20] for the learning process of a machine because it helps to get effective results as soon as possible. C4.5 also known as J48 uses a pruning method for building a tree. This algorithm continues to be a recursive process until the expected results are found. It provides good accuracy and flexibility. This formula is made available from the following equations in [21].

$$E = \sum_{i=1}^{K} P_i \, log_2 \, P_i \quad (11)$$

From equation (1), K defines the number of classes of expected properties, Pi defines the number of appearance of class, i is divided by the sum of instances. Fig. 6 shows the basic structure of a conditioned C4.5 decision tree.

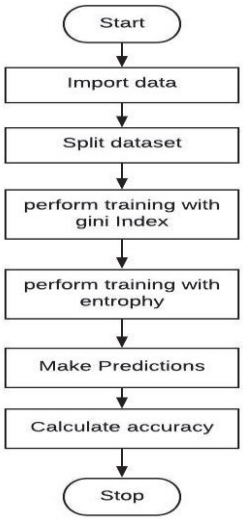

Fig. 6. Decision tree flowchart

F. *Random Forest:*

Random forest is considered as an ensemble machine learning method for classification, regression, and probability[22]. A large number of decision trees are used to create a random f rest algorithm. The random forest [23] is able to find missing values from a large number of datasets and it is capable of providing a more accurate value. The working procedure of random forest is discussed in Fig. 7.

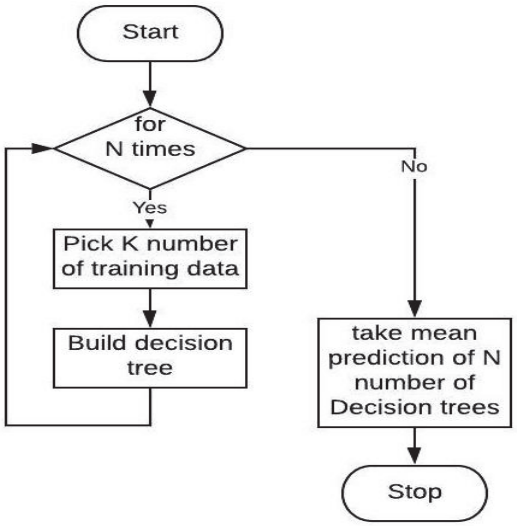

Fig. 7. Working process of the Random forest

G. *K-Nearest Neighbor:*

The K-Nearest Neighbor is known as one of the most used algorithms in machine learning due to its versatility[24]. 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. Euclidean distance equation is: [25].

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \qquad (12)$$

The construction of the Knn algorithm is shown below in the figure 8.

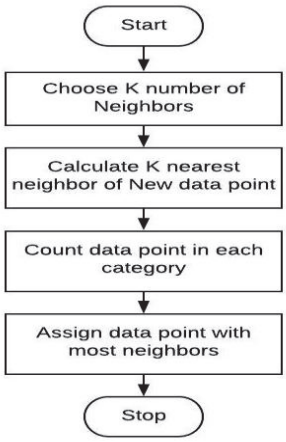

Fig. 8. Construction of the Random forest

V. RESULTS AND DISCUSSION

A. *Outputs of Data pre-processing:*

The data cleaning strategy is considered to be the top strategy for creating any model. Filtering with changes is applied to realize the data. The correlation between different predictors is shown in the heat map below in fig. 9.

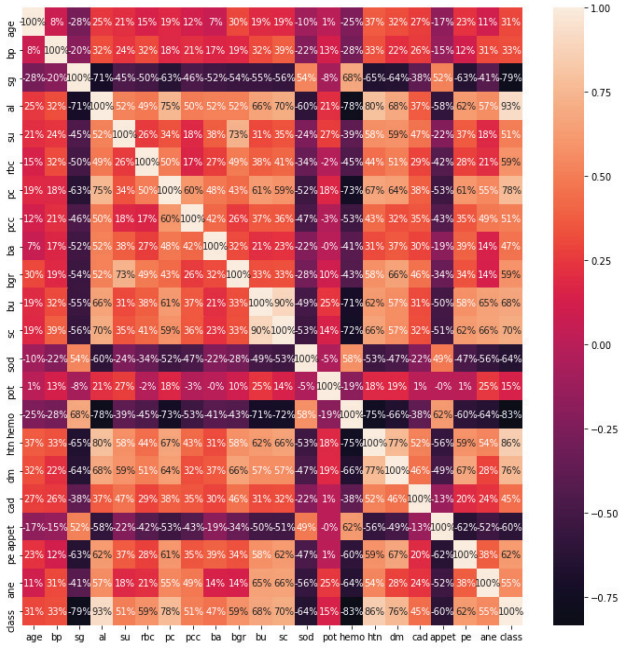

Fig. 9. Correlation Between Different Predictor Using Heat – Map

Following the dataset's effective evaluation strategy, a large amount of data has been shared in training and testing. To identify whether a patient has KD or not, four methods of classification and regression techniques were used. Performance metrics such as confusion matrix, accuracy, f1-score, recall (Sensitivity), and specification are used to justify different algorithm methods. Positive classification occurs when a person has symptoms of kidney disease and negative classification occurs when a person does not have kidney disease (KD), The accuracy rate of the prediction model is too high for both training and testing data.

The highest outcomes achieve from Logistic Regression, and Random Forest, The same results are obtained from sensitivity, specificity values as well as precision.

B. *Experimental Outputs Among Different Methods:*

All of the techniques such as Confusion Matrix, Accuracy, Sensitivity, Specificity, Precision, Negative predictive value, False Positive Rate, False Discovery Rate, False Negative Rate, F1 Score, Standard Deviation, Mean Absolute Error, Mean Squared Error and Root Mean Squared Error are studied to make a kidney disease model in below.

Logistic regression and Random Forest provides 100% accuracy with a 1.0 Sensitivity score and 1.0 recall score respectively. This accuracy is obtained due to over fitting. The second-highest score is obtained from decision tree with accuracy (97.91%) and recall (0.9722) value to predict kidney disease From KNN, the poorest score of 95.83% accuracy and 0.9459 recall (sensitivity) score is gained. The matrix outputs are shown in Table 1.

TABLE I. PERFORMACE MEASUREMENT CRITERIA

Dimension	Value			
	Logistic regression	decision tree	Random Forest	K-Nearest Neighbors
Confusion Matrix	[35 0] [0 13]	[35 0] [1 12]	[35 0] [0 13]	[35 0] [2 11]
Accuracy	100%	97.91%	100%	95.83%
Sensitivity	1.0	0.9722	1.0	0.9459
Specificity	1.0	1.0	1.0	1.0
Precision	1.0	0.98	1.0	0.96
Negative predictive value	1.0	0.9230	1.0	0.8461
False Positive Rate	0.0	0.0	0.0	0.0
False Discovery Rate	0.0	0.0	0.0	0.0
False Negative Rate	0.0	0.0277	0.0	0.0540
F1 Score	1.0	0.98	1.0	0.96
Standard Deviation	0.0000	0.1705	0.0000	0.2179
Mean Absolute Error	0.0000	0.0208	0.0000	0.0417
Mean Squared Error	0.0000	0.0208	0.0000	0.0417
Root Mean Squared Error	0.0000	0.1443	0.0000	0.2041

A successful detection rate between perfect machine learning techniques is presented in Fig. 10 to show visual satisfaction.

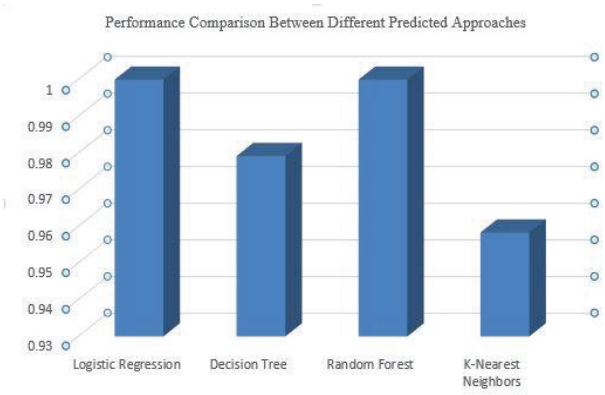

Fig. 10. Performance comparison between various discussed machine learning tools.

C. *Execution Time Measurement over the applied Models:*

The prediction rate is totally dependent on the dataset along with the model preprocessing technique. Besides, time takes a prominent role compared to others. As we know from Fig. 10 that the highest predictable accuracy comes from two different techniques where the ratio of run time is very low for Logistic Regression (LR). Alternatively, Random Forest takes the highest time to achieve a predictable score that has been clearly illustrated in fig. 11. Other popular algorithms generate lower time periods to find out the represented outcomes.

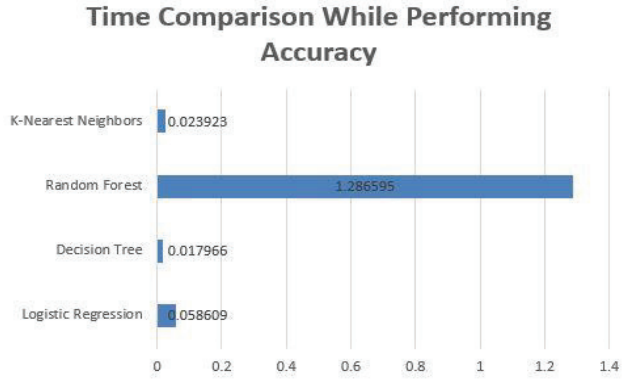

Fig. 11. A Predictable Time Comparison Occurred on Performed Algorithms.

D. *Error Rate Among Various approaches:*

An error rate is an approach that helps us to take an accurate decision of the working methods. This also helps us to know about the given model predicted accuracy. The highest outcome has been achieved from four renewable techniques such as Logistic Regression, Random Forest, Support Vector Machine, and Gradient Boosting; however, the lowest result is gained from K-Nearest Neighbors.

As Logistic Regression provides the highest accuracy such as 100% with 1.0 recall score same as random forest, there has been no error rate found. Decision Tree Classifier generates the mid-lowest accuracy of 97.91% with 1.0 specificity score. The K- nearest neighbors provide the lowest accuracy among them. The error rate is very high because of its poor accuracy. Fig. 12. explains the outcomes.

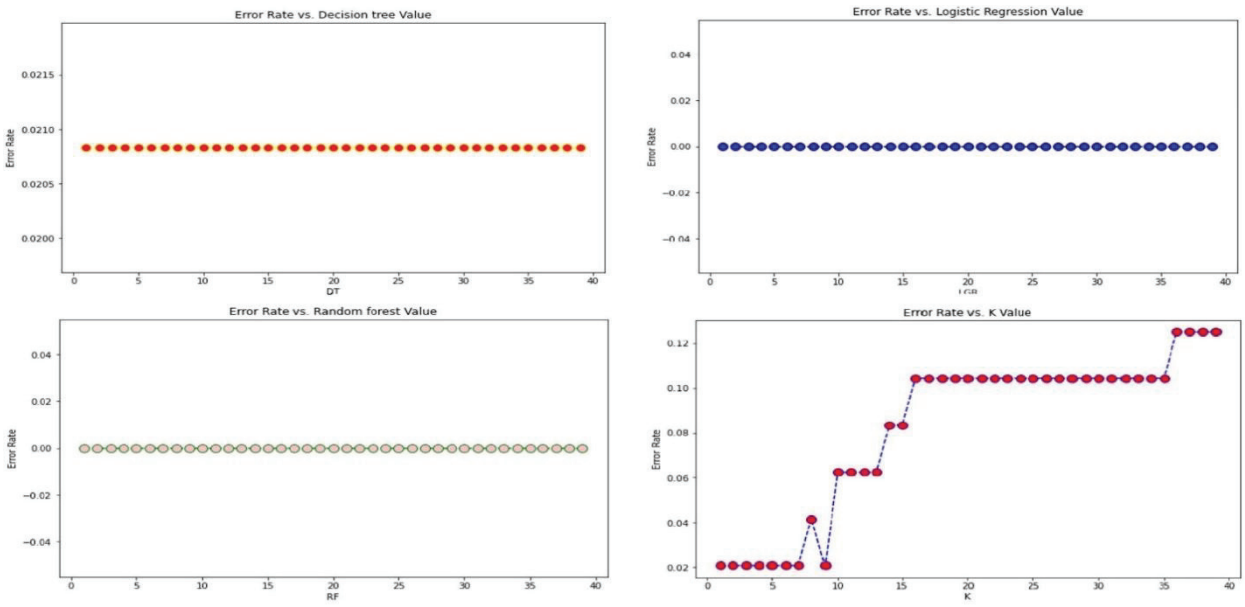

Fig. 12. The error rate of the models

E. Results to Detect KD Using ROC and AUC Curves::

The diagnostic talents of the classifiers are usually confused using the confusion matrix and the receiver working property (ROC) curve [26]. The matrix of confusion is also called the contingency matrix in the study of machine learning. True positive (TP) occurs when classifiers identify them effectively [27][28]; However, False Positive (FP) occurs when classifiers cannot identify them efficiently. To understand the overall performance of the model, AUC and ROC matrix devices were used to test the learning performance. The ROC curve is determined because the receiver is the working characteristic curve that lies below the AUC and ROC curves. If the AUC rating is excessive, the performance of the version should be excessive [29] and vice versa. Logistics regression, random forest, K-Nearest neighbor, provide the best ratings on each ROC and AUC diagram.

Logistic Regression gives the highest score in both ROC and AUC curves whereas Decision Tree Classifier and K-Nearest Neighbors gives the most reduced score. The score of Random forest affords the very best rating. ROC and AUC curves of the Classifier is illustrated in Fig. 13.

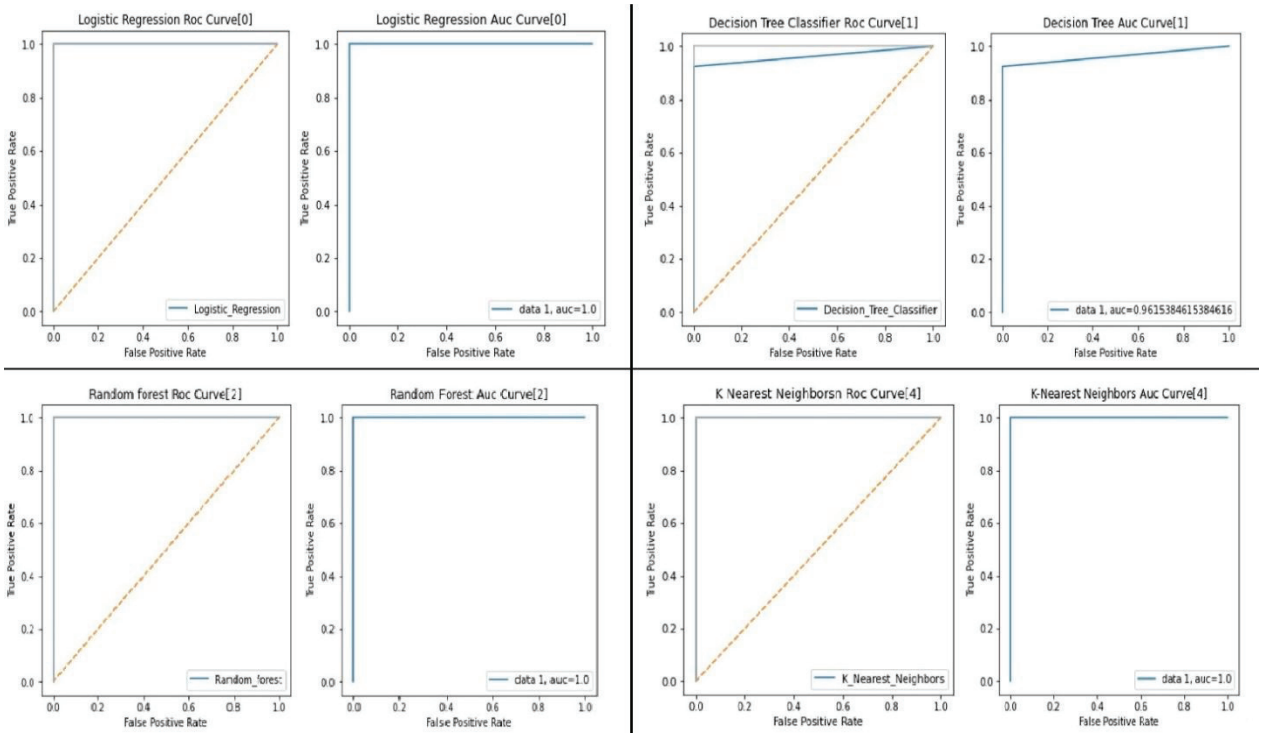

Fig. 13. ROC and AUC curves of classifers

VI. Conclusions

In this paper, we took four different machine learning algorithms in order to get an accurate prediction rate over the presented dataset. Comparing all introduced approaches, the successful outcomes have been obtained from four perfect machine learning techniques such as Logistic Regression, and Random Forest. These models successfully generate a 100% accuracy rate on the recognized dataset. On the other hand, the lowest predictable score comes from K (95.83%) which was a pretty low score among other exhibited classifiers. It is also seen that the Random forest takes more time to predict result and the best rating in ROC curve. Furthermore, the KNN algorithm has higher error rate compared to others. Since an accurate rate of prediction is undoubtedly dependent on the pre-processed technique, the techniques of the pre-processed model must handle carefully to achieve more correct results.

Acknowledgment

The authors are grateful and pleased to all the researchers in this research study.

References

[1] F.M. Javed Mehedi Shamrat, Md. Asaduzzaman, A.K.M. Sazzadur Rahman, Raja Tariqul Hasan Tusher, Zarrin Tasnim "A Comparative Analysis Of Parkinson Disease Prediction Using Machine Learning Approaches" International Journal of Scientific & Technology Research, Volume 8, Issue 11, November 2019, ISSN: 2277-8616, pp: 2576-2580.

[2] A.K.M Sazzadur Rahman, F. M. Javed Mehedi Shamrat, Zarrin Tasnim, Joy Roy, Syed Akhter Hossain "A Comparative Study On Liver Disease Prediction Using Supervised Machine Learning Algorithms" International Journal of Scientific & Technology Research, Volume 8, Issue 11, November 2019, ISSN: 2277-8616, pp: 419-422.

[3] F. M. Javed Mehedi Shamrat, Md. Abu Raihan, A.K.M. Sazzadur Rahman, Imran Mahmud, Rozina Akter, "An Analysis on Breast Disease Prediction Using Machine Learning Approaches" International Journal of Scientific & Technology Research, Volume 9, Issue 02, February 2020, ISSN: 2277-8616, pp: 2450-2455.

[4] Zaki, Z., Shah, M. A., Wakil, K., Sher, F., : Logistic regression based human activities recognition. In: Journal of mechanics of continua and mathematical sciences. Vol.-15, No.-4, April (2020) pp 228-246

[5] Reddan,T., Corness, J., Harden, F., Mengersen, K. : Analysis of the predictive value of clinical and sonographic variables in children with suspected acute appendicitis using decision tree algorithms. In: Australasian Sonographers Association, Sonography 5 pp.157–163 (2018)

[6] Ong J, Liu X, Rajarethinam J, Kok SY,Liang S, Tang CS, et al. (2018) Mapping denguerisk in Singapore using Random Forest. PLoS Neglected Tropical Diseases 12(6): e0006587. https://doi.org/10.1371/journal.pntd.0006587

[7] Suhartono, Nguyen, P. T., Shankar, K., Hashim, W., Maseleno A. : Brain Tumor Segmentation and Classification using KNN Algorithm. In: International Journal of Engineering and Advanced Technology, ISSN: 2249 – 8958, Volume-8, Issue-6S, pp 706-711 (2019)

[8] "UCI machine learning repository: Early stage of Chronic kidney disease dataset, https://archive.ics.uci.edu/ml/datasets/Chronic KidneyDisease, (Accessed on 05/06/2020).

[9] P. Yildirim, "Chronic Kidney Disease Prediction on Imbalanced Data by Multilayer Perceptron: Chronic Kidney Disease Prediction," 2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC), Turin, pp. 193-198, 2017.

[10] "Smarter ways to encode categorical data for machine learning,"https://towardsdatascience.com/smarter-ways-to-encode-categoricaldata-for-machine-learning-part-1-of-3-6dca2f71b159, (Accessed on 05/06/2020).

[11] Zheng, Lijuan, H. Wang, and S. Gao, "Sentimental feature selection for sentiment analysis of Chinese online reviews", International Journal of Machine Learning and Cybernetics, Vol. 6, March, 2015.

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

[13] R. A. Alassaf et al., "Preemptive Diagnosis of Chronic Kidney Disease Using Machine Learning Techniques," International Conference on Innovations in Information Technology (IIT), Al Ain, pp. 99-104, 2018.

[14] S. Manikandan, "Heart attack prediction system," International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), Chennai, pp. 817-820, 2017.

[15] A. A. Imran, M. N. Amin and F. T. Johora, "Classification of Chronic Kidney Disease using Logistic Regression, Feedforward Neural Network and Wide & Deep Learning," International Conference on Innovation in Engineering and Technology (ICIET), Dhaka, Bangladesh, pp. 1-6, 2018.

[16] S. Bharati, M. A. Rahman and P. Podder, "Breast Cancer Prediction Applying Different Classification Algorithm with Comparative Analysis using WEKA," 2018 4th International Conference on Electrical Engineering and Information & Communication Technology (iCEEiCT), Dhaka, Bangladesh, pp. 581-584, 2018.

[17] F. Muratore, L. Boiardi and A. Cavazza, et al. "Correlations between histopathological findings and clinical manifestations in biopsy-proven giant cell arteritis", Journal of Autoimmunity. Vol. 69, pp. 94-101, 2016.

[18] L. Nguyen, A. B. Tosun, J. L. Fine, D. L. Taylor, and S. C. Chennubhotla, "Architectural Patterns For Differential Diagnosis Of Proliferative Breast Lesions From Histopathological Images," IN IEEE 14th International Symposium on Biomedical Imaging (ISBI), pp. 152–155, 2017.

[19] P. S. Kohli and S. Arora, "Application of Machine Learning in Disease Prediction," 2018 4th International Conference on Computing Communication and Automation (ICCCA), Greater Noida, India, pp. 1-4, 2018.

[20] P. Ghosh, M. Z. Hasan and M. I. Jabiullah, "A Comparative Study of Machine Learning Approaches on Dataset to Predicting Cancer Outcome," Bangladesh Electronic Society, Vol. 18 (1–3), pp. 01–05, 2018.

[21] A. K. Pandey, P. Pandey, K. L. Jaiswal, A. K. Sen," A Heart Disease Prediction Model using Decision Tree," IOSR Journal of Computer Engineering (IOSR-JCE), Vol. 12, pp 83-86 , 2013.

[22] J. Han, M. Kamber and J. Pei, "Data Mining Concepts and Techniques (3rd)," The Morgan Kaufmann Series in Data Management Systems, 2011.

[23] 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, pp. 679-684, 2019.

[24] H. Alasker, S. Alharkan, W. Alharkan, A. Zaki and L. S. Riza, "Detection of kidney disease using various intelligent classifiers," 2017 3rd International Conference on Science in Information Technology (ICSITech), Bandung, pp. 681-684, 2017.

[25] S. V. J. Jaikrishnan, O. Chantarakasemchit and P. Meesad, "A Breakup Machine Learning Approach for Breast Cancer Prediction," 11th International Conference on Information Technology and Electrical Engineering (ICITEE), Pattaya, Thailand, pp. 1-6, 2019.

[26] S. Uddin, A. Khan, M. E. Hossain, M. A. Moni, "Comparing different supervised machine learning algorithms for disease prediction," BMC Medical Informatics and Decision Making, Vol. 19, 21 Dec, 2019.

[27] Pronab Ghosh, Asif Karim, Syeda Tanjila Atik, Saima Afrin, mohd Saifuzzaman," Expert Model of Cancer Disease Using Supervised Algorithms with a LASSO Feature Selection Approach," International Journal of Electrical and Computer Engineering,Volume 11, Issue 3, 2020

[28] FA Ahmed, MS Islam, A Karim, N Neehal : Shot-Net: A Convolutional Neural Network for Classifying Different Cricket Shots. In: Springer: Communications in Computer and Information Science (CCIS), vol. 1035, 2019

[29] A Karim, S Azam, B Shanmugam, K Kannoorpatti, M Alazab : A Comprehensive Survey for Intelligent Spam Email Detection. In: IEEE Access 7 (1), 168261-168295, (2019)