*Uluslararası İleri Doğa Bilimleri ve Mühendislik Araştırmaları Dergisi*
*Sayı 7, S. 225-233, 10, 2023*

***Araştırma Makalesi***

IJANSER
https://alls-academy.com/index.php/ijanser
ISSN: 2980-0811

*International Journal of Advanced Natural Sciences and Engineering Researches*
*Volume 7, pp. 225-233, 10, 2023*

***Research Article***

# Enhancing Chronic Kidney Disease Diagnosis using Machine Learning Classifiers: A Comparative Analysis

Abdul Majid[1], Muhammad Ismail[1*], Fazal Muhammad[1], Jamal Hussain Arman[1], Bilal Khan[2]

[1]*Department of Electrical Engineering, University of Engineering and Technology Mardan, 23200 Mardan, Pakistan*
[2]*Department of Computer Science, City University of Science and Information Technology, 25000, Peshawar, Pakistan*

[*]*(m.ismail012018@gmail.com) Email of the corresponding author*

*(Received: 02 November 2023, Accepted: 13 November 2023)*

(2nd International Conference on Contemporary Academic Research ICCAR 2023, November 4-5, 2023)

**ATIF/REFERENCE:** Majid, A., Ismail, M., Muhammad, F., Arman, J. H. & Khan, B. (2023). Enhancing Chronic Kidney Disease Diagnosis using Machine Learning Classifiers: A Comparative Analysis. *International Journal of Advanced Natural Sciences and Engineering Researches*, 7(10), 225-233.

***Abstract*** **–**One of the most common and dangerous illnesses affecting people on a global scale, chronic kidney disease (CKD), does not manifest itself until the kidneys of a particular person have sustained irreparable harm. The progression of CKD is linked to many serious side effects, such as an increased risk of different diseases, kidney failure, nerve harm, pregnancy problems, anemia, and hyperlipidemia. This illness claims the lives of millions of individuals each year. Since there are no significant symptoms that can be used as a benchmark to identify CKD, diagnosing the condition might be difficult. Occasionally, data may be interpreted wrongly when the diagnosis is persistent. To diagnose CKD in patients, this study employs a machine learning classifier. Six machine learning (ML) techniques are used in this study, including Random Forest (RF), Random Tree (RT), Decision Table (DTa), Decision Tree (DTr), Naïve Bayes (NB), and Hoeffding Tree and multiple performance metrics are considered such as accuracy, TPR, FPR, recall and mean absolute error (MAE). To select the most accurate classifier for predicting CKD, these predictive models are created using a dataset on chronic kidney disease containing 279 attributes acquired from Kaggle. Our objective is to ease the introduction of machine learning techniques for precisely detecting CKD by learning from dataset attribute reports. The main contribution of the research is an ML-based model for diagnosing chronic renal disease that outperforms common diagnosing techniques and reaches the highest predicted accuracy. This study also contrasted how well each model performed. We were able to predict this disease with the Random Forest model more accurately than ever before, at a 76.23% accuracy level.

*Keywords – Chronic Kidney Disease, Machine Learning, Random Forest, Performance Metrics*

## I. INTRODUCTION

Chronic kidney disease (CKD) is a condition that is extremely uncomfortable and eventually fatal and affects 10% of people worldwide today. Growing to be the 18th most fatal disease in 2010 [1], it not only signifies the onset of renal failure but also encourages the development of further diseases over the course of the patient's lifetime if treatment is not given or taken as directed. This illness is made all the deadlier by the fact that it cannot be identified until the kidneys have already sustained irreparable harm. By the time a patient realizes a health issue, it has already taken a long time to get him tested, diagnose a result that might not be accurate, give

him medication based on the stage of CKD he might be experiencing, and provide him all the treatment he needs to be alive. Choosing the appropriate amount of erythropoietin dosage to administer to thousands of patients within a month is one of the difficulties associated with CKD. DARWIN is an intelligent software decision support system that is used to track how much erythropoietin should be provided to CKD patients [2].

One of the main reasons chronic kidney disease has become one of the worst diseases in the world is that there isn't a single, broadly applicable indicator that can be used to categorically discriminate between healthy and diseased people. This makes it more challenging for medical professionals to detect this condition promptly and precisely, which results in erroneous disease forecasts [3]. The principle of supervised machine learning may be used to help the model recognize various data patterns [4]. To increase test accuracy, nevertheless, a reliable categorization model that is unaffected by changing circumstances is needed. By adjusting hyperparameters and giving sufficient input data, the trained artificial neural network (ANN) may achieve significant test accuracy in discriminating CKD patients from the rest [5]. In [6,7], the authors claim that the application of neural networks in medicine delivery can help individuals with chronic renal failure. Since neural networks (NN) have been developed over time and are now capable of tasks like pattern recognition, prediction, classification, etc., they seem to be a good option for the purpose of diagnosing CKD.

Artificial intelligence (AI) in medicine may be developed using sizable, actual clinical data sets. It is difficult for people to directly study these enormous data sets due to the time and caution needed to prevent human mistakes, along with the ability to completely extract the insights or information. It is patently clear that in some situations, artificial intelligence systems perform orders of magnitude better than humans. Research on AI and renal illness has just begun. Four main areas— prognosis assessment, early warning systems, treatment planning, and diagnostic assistance, are the main subjects of current research on AI's potential impact on renal illness and CKD.

Due to the high rates of morbidity and mortality that both acute and chronic renal illnesses generate, as well as the significant economic effect, renal disease is a significant public health and medical burden around the world. The appearance, development, and response to treatment of patients with renal illness fluctuate greatly. For more accurate phenotypic and outcome prediction in renal disease, AI can provide information on precision therapy. Medicine is the study of the kidneys and their problems in children and adults. The nephrologists are in charge of treating and diagnosing renal failure. The kidney is necessary for maintaining the body's natural balance of electrolytes and water [8].

The most frequent methods for identifying chronic kidney disease are screenings that involve a blood chemical profile and urine tests or diagnosing the condition after a separate operation. Less common symptoms include flank pain, severe haematuria, "foamy pee" nocturia, and reduced urine production. People with severe CKD may have dyspnoea, sudden weight loss, exhaustion, vomiting, decreased appetite, metallic taste, nausea, peripheral edema, pruritus, changes in mental status, and more.

## II. METHODOLOGY

New techniques for creating a prediction model, which previously relied on conventional statistics, became available with the introduction of the big data age. Artificial intelligence (AI) is a subset that includes machine learning (ML), which enables the computer to carry out a certain activity without explicit instructions. The ML algorithm can be trained to recognize the underlying patterns of the sample data and forecast the new data based on the learned knowledge when used in predictive modeling [9, 10].

The main objective of the proof-of-concept study presented in this work was to develop ML models for predicting the risk of ESKD on a dataset of Chinese CKD cases. The baseline features and common blood tests were used as the basis for training and testing the machine-learning models. The findings of this study indicate that ML models are not only capable of carrying out this crucial clinical duty but also have the potential to advance customized medicine.

This section talks about the CKD dataset, 10-fold cross-validation (10FCV), performance evaluation,

and comparative analytic methodologies. This study uses an updated dataset retrieved from the UCI repository and focuses on CKD. Following the description of the dataset, the 10FCV method is used for training, and various testing techniques are used. Finally, the results of the proposed model and those of the other models are compared using performance metrics such as accuracy, true positive rate (TPR), false positive rate (FPR), recall, and mean absolute error (MAE). Fig.1. depicts the whole experimental setup.

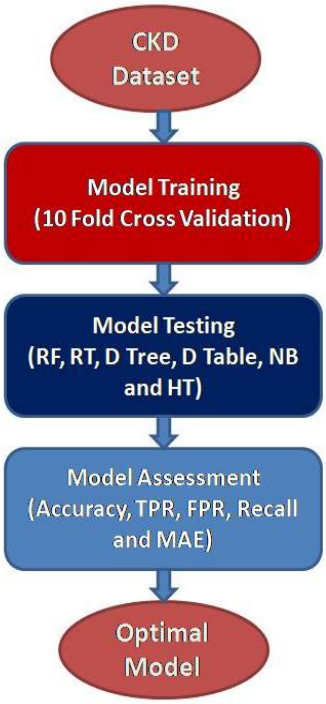


Fig.1. Flow chart of proposed study

## III. DATASET AND ML MODELS

### *A. CKD Dataset*

The database for arrhythmia, UCI Respiratory, provided the data for the study. This dataset was selected due to how poorly it has been applied in previous research papers. There are 279 attributes in this dataset, 206 of which have linear values, and the remaining ones are nominal [11].

### *B. Dataset Description*

Twenty-Five features are selected QRS Interval, QT Interval, T Interval, T, Heart Rate, Channel DII, Channel AVF, 2 Channel V1, 5 Channel V1, Channel V2, 3 Channel V2, Channel V3, 2 Channel V3, 12 Channel AVR, 19 Channel AVR, 13 Channel AVF, 19 Channel AVF, 14 Channel V1, 16 Channel V1, 19 ChannelV3, 20 Channel V3, 19 Channel V5, 19 Channel V6, 21 Channel V6, Class [12].

Table.1. Dataset attributes

| S.No | Attributes | Description |
|---|---|---|
| 1 | QRS Interval | Average of QRS duration in msec., linear |
| 2 | QT Interval | Average duration between onset of Q and offset of T waves in msec., linear |
| 3 | T Interval | Average duration of T wave in msec, linear |
| 4 | T | Repolarization of the Ventricle |
| 5 | Heart Rate | Number of heart beats per minute, linear |
| 6 | Channel DII | Average width, in msec., of: linear |
| 7 | Channel AVF | Augmented Vector |
| 8 | 2 Channel V1 | Right Voltage (RV) |
| 9 | 5 Channel V1 | Right Voltage (RV) |
| 10 | Channel V2 | Right Voltage (RV) |
| 11 | 3 Channel V2 | Right Voltage (RV) |
| 12 | Channel V3 | Septum |
| 13 | T | Repolarization of the Ventricle |
| 14 | 2 Channel V3 | Septum |
| 15 | 12 Channel AVR | Augmented Vector Right |
| 16 | 19 Channel AVR | Augmented Vector Right |
| 17 | 13 Channel AVF | Augmented Vector Foot |
| 18 | 19 Channel AVF | Augmented Vector Foot |
| 19 | 14 Channel V1 | Right Voltage (RV) |

| | | |
|---|---|---|
| 20 | 16 Channel V1 | Right Voltage (RV) |
| 21 | 19 Channel V3 | Septum |
| 22 | 20 Channel V3 | Septum |
| 23 | 19 Channel V5 | L side of the heart |
| 24 | 19 Channel V6 | L side of the heart |
| 25 | 21 Channel V6 | L side of the heart |

AI is having a stronger impact on healthcare and is progressively changing the ways in which clinicians approach problem-solving because of the widespread usage of electronic health data and recent advances in machine learning research. When training an ML model, data-driven methods are typically used rather than theory-driven approaches, which force the model to start with a predetermined hypothesis based on prior information. When the predictions and the actual results are compared, the model iteratively improves its performance on a training set by adjusting model parameters to close the difference between the two.

In the field of nephrology, machine learning has shown promise in predicting acute kidney injury or time to allograft loss from clinical features, recognizing particular patterns in pathology slides, selecting the best prescription for dialysis, and mining text in the electronic health record to find particular cases. Recent research used machine learning (ML) approaches to forecast the onset of CKD. These models were developed to predict future eGFR values, calculate the risk of short-term mortality after dialysis, or assess the protein concentrations in the urine for 24 hours [13].

### *C. Training and Testing*

To use the dataset in the experiment, we divided the preprocessed data into training and testing sets. During the training phase, a 10FCV model receives the training set's processed data. Cross validation, this comprises withholding a portion of the data, training the model with the remaining data, and testing it with the withheld data, is a useful technique for assessing the model's performance [14]. Fig.2. shows the flow chart of cross-validation.

The classifiers are evaluated using the specified testing set during the testing phase. In the testing set, the classifier generates either a normal or an attack label for each instance of data. In this stage, various ML approaches were used.

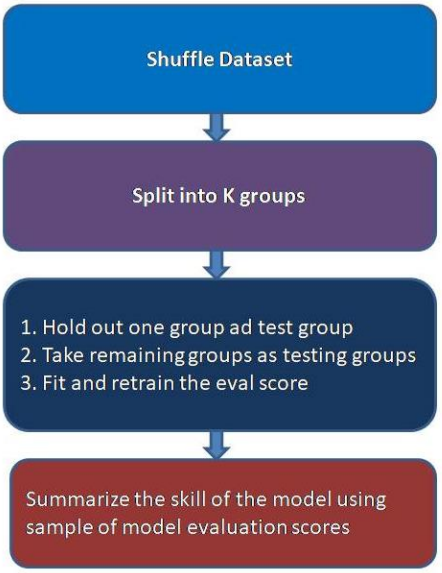


Fig. 2. Cross Validation Flowchart

### *D. Proposed ML Teachniques for Model Training*

In this study we used Random Forest (RF) based model to train our model. This technique discussed below.

- *Random Forest*

A variety of decision trees are used by the Random Forest (RF) and are applied to different datasets. The RF classifier then combines the decision tree findings to significantly increase prediction accuracy. RF is used to solve regression and classification problems. To reduce the risk of overfitting, it integrates a number of decision trees. Since RF can handle different features, feature scaling is not necessary.

RF outperforms other strategies in terms of classification accuracy by overcoming the issues of overfitting and volatility. Depending on the node for which the optimum split point technique is derived, this prediction mechanism selects dependent qualities from the entire collection of attributes in

the dataset. As a result of this splitting procedure, multiple trees are created when all characteristics have been separated into nodes. Then, all of the produced trees are integrated to create the RF model.

The Pseudo Code of RF are:

1. Select "k" features randomly from a pool of "m" features.
2. Locate the node "d" by the better-fragmented feature among the "k" features.
3. Split the node into child nodes using the tiniest fragments.
4. Repeat steps 1 through 3 until the "l" number of nodes is understood.
5. Create the forest by repeating steps 1 through 4 n times to build n trees.

RF is a supervised machine learning method that uses labeled data. Various analytical techniques have been used in earlier study. CKD dataset with a large number of attributes was employed in this work, and the overfitting issue is typically a concern with big datasets. We employed an RF-based classification model, which considerably decreases the overfitting problems [15].

## IV. THE METHODS USED FOR COMPARATIVE ANALYSIS ARE

To compare the outcomes of our suggested models with those of other ML approaches, we have also used some of them. Following are some methods for comparative analysis:

### *A. Random Tree*

The RC method is used to build a number of fundamental classifiers using a range of seed values selected at random. Merging the findings of classifiers that provide diverse values and reducing the likelihood of error, improves the outcome. The procedure of producing several base classifiers using various random number values yields the final classification result by averaging the predictions given by each basis classifier. The most effective method for handling the values of binary, numeric, nominal, and missing classes is RC, a class for building an ensemble of randomizable base classifiers [16].

### *B. Decision Tree*

The decision tree algorithm known as J48 (C4.5), which is an extension of Quinlan's previous ID3 Algorithm and is known to have a respectable accuracy rate in bio-medical applications, is the most often used decision tree algorithm. It is capable of handling both categorical and numerical data. Another name for it is statistical classifier. It handles both noise and missing values and is simple to build. J48's performance is also subpar for a little training set.

The J48 algorithm employed in this study generates output based on the following steps:

6. Select the dataset to be used as the rule's input. The J48 method splits category attributes in the same way as the ID3 algorithm.
7. Determine each feature's Normalized information gain.
8. The best trait is determined to be the one that provides the most information gain. To build a decision tree, the attribute with the highest gain is chosen as the root node.
9. To calculate the information gain for each attribute and add it as a child node, repeat steps 1 through 3 until a stop condition is reached.

Based on the existence or nonexistence of DM, age, hypertension, and gender, the J48 algorithm divides the patient's situation into four classes: low, moderate, high, and very high risk during the CKD risk analysis. In cases of moderate risk, high risk, or extremely high risk, the sub-system presents the categorization findings into a CDA document along with the decision tree graphic and the patient's general information [17].

### *C. Decision Table*

Machine learning can effectively express and organize complex decision reasoning using decision tables. They offer a methodical technique to specify rules and conditions based on the characteristics of the input and choose the appropriate output or course of action. It's crucial to remember that decision tables are frequently employed for models that are clear and easy to understand. Based on cost-sensitive learning, we multiply, where is the percentage of the total number of positive and negative training samples to increase the number of

negative samples in each column of the uneven decision table (DT). As a result, each DT column's decision weight for negative samples is altered. In this method, the problem of uneven categorization is handled. The decision tree is one of the most well-known classification and prediction techniques in ML. In the tree-like structure known as a decision tree, an internal node reflects qualities, a branch indicates the result, and leaves represent a class label. These algorithms generate decision rules that predict the outcomes of fictitious test cases [18].

These algorithms offer high precision and enhanced interpretation. The decision tree can be utilized with discrete and continuous data. Complex decision rules are illustrated by a DT, which is a tabular shape with rows and columns.

*D. Naive Bayes*

The Naive Bayes (NB) algorithm uses conditional probability to forecast the class when some historical data is provided. Whenever the data contains more independent variables, NB, which demonstrates outstanding accuracy, is the best classification technique for categorical data. Text classification is the industry that Naive Bayes mostly targets. It is mostly employed for classification and clustering purposes. Conditional probability is a key component of the naive Bayes architecture. Depending on their likelihood of occurring, it generates trees. Bayesian networks are another name for these trees. It is crucial to keep in mind that when we discuss the Naive Bayes classifier, we define it as a simple and straightforward classifier based on probability, with its bases being the assumption of strong (naive) individuality and the Bayes' theorem (which is a fundamental component of Bayesian statistics). The most effective approach to describe the underlying model, which is probabilistic in nature, is an "independent feature model." However, this limiting individuality assumption seems to be mostly false when taking into account practicalities and purposes involving a particularly realistic nature.

Thus, we use the term "naive," and despite the aforementioned issue, the algorithm is extremely effective and a quick learner, making it appropriate for supervised classification cases. In other words, one of the main benefits of the Naive Bayes classifier is the efficient approximation of the parameters, which are typically the mean and variance, even when only a small percentage of the training data is used. Due to the basic premise of individual variables, it is not necessary to determine the entire variance matrix for each class, only the variances related to the variables [19].

*E. Hoeffding Tree*

The Hoeffding tree is a decision tree type that is intended for incremental learning or streaming data. It is renowned for its effectiveness in producing trees that are comparable to the types that would be produced by training on the complete dataset simultaneously in a non-streaming manner. In this research, we compared our recommended strategy to another incremental learning method built on trees. We opt for the well-known incremental decision tree method known as the Hoeffding tree (HT) algorithm. Massive amounts of incremental data can be used to train the HT technique. It makes use of the fact that selecting the ideal dataset-splitting attribute frequently only requires a small quantity of data. Thus, one common application of this technique is the processing of incremental data [20].

## V. PERFPRMANCE EVALUATION METRICS

Among the performance indicators utilized in this work to compare outcomes and assess performance were accuracy, TPR, FPR, MAE, and recall. Only those aspects are selected that are pertinent and essential to kidney diseases. The suggested classifier's performance is then evaluated using performance assessment measures. The number of predictions that were successfully classified as risk on all predictions made using the available dataset is known as accuracy, which is the most common technique of performance measurement [21]. Confusion matrix can be used to determine evaluation metrics outcomes. Confusion Matrix, often known as an error matrix, is a performance evaluation chart used for classification issues in machine learning. It essentially shows how frequently predictions made by the categorization model were incorrect [22]. Table.2. displays a confusion matrix with four parameters, and Table.3.

displays the mathematical formulas for our performance measures.

Table.2. Representation of a confusion matrix

| **Actual Class** | **Predicted Class** | |
|---|---|---|
| | **Positive** | **Negative** |
| Positive | TP | FN |
| Negative | FP | TN |

Table.3. Performance assessment metrics mathematical forms

| **Performance Metrics** | **Mathematical Form** |
|---|---|
| Accuracy | (TP+TN)/(TP+TN+FN+FP) |
| TPR | (TP)/(TP+ FN) |
| FPR | (FP)/( TN+ FP) |
| Recall | (TP)/(TP+ FN) |
| MAE | $\frac{1}{n}\sum_{i=1}^{n} y - \bar{y}$ |

## VI. RESULT AND DISCUSSION

Five performance assessment metrics are used in this section to show the outcomes of the proposed and other ML models. The findings of the used models are shown in Table.4. and additional figures display comparative analyses of all used models, which are covered below.

Table.4. Applied Techniques Results

| **Techniques** | **Accuracy** | **TPR** | **FPR** | **Recall** | **MAE** |
|---|---|---|---|---|---|
| RF | 76.23 | 0.763 | 0.181 | 0.763 | 0.0441 |
| RT | 59.51 | 0.595 | 0.187 | 0.595 | 0.0626 |
| D Tree | 64.16 | 0.692 | 0.148 | 0.692 | 0.055 |
| D Table | 65.0442 | 0.65 | 0.232 | 0.65 | 0.0839 |
| NB | 69.69 | 0.697 | 0.163 | 0.697 | 0.0475 |
| HF | 54.2035 | 0.542 | 0.542 | 0.542 | 0.1051 |

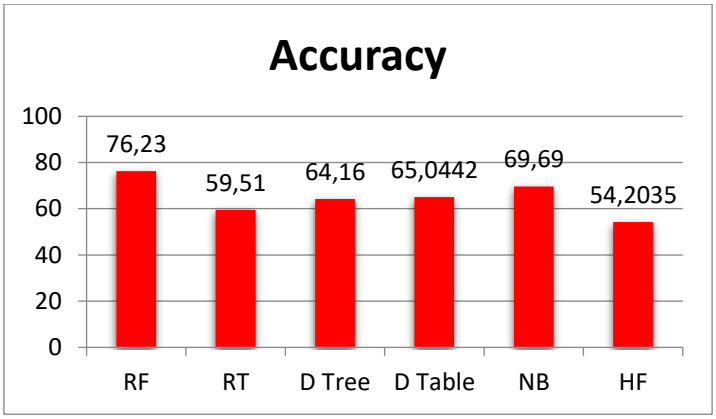


Fig.3. Accuracy values of the applied techniques

Fig.3. shows the comparison of random forest, random tree, decision tree, decision table, naïve bayes and hoeffding tree in terms of accuracy in which RF achieves highest accuracy of 76.23% among all other techniques.

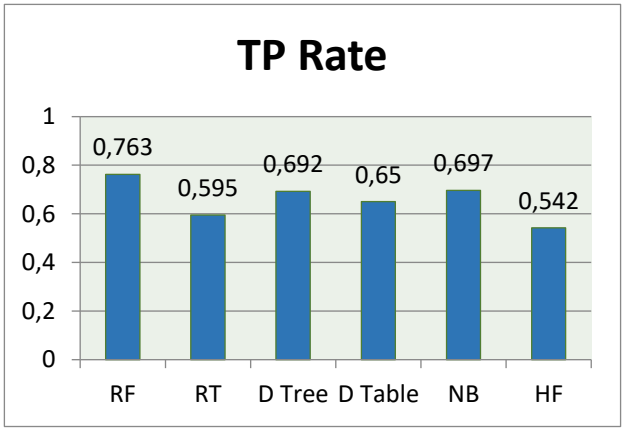


Fig .4. True positive rate of the applied techniques

Fig.4. shows the comparison of random forest, random tree, decision tree, decision table, naïve bayes and hoeffding tree in terms of true positive rate in which RF achieves highest TPR of 0.763 among all other techniques.

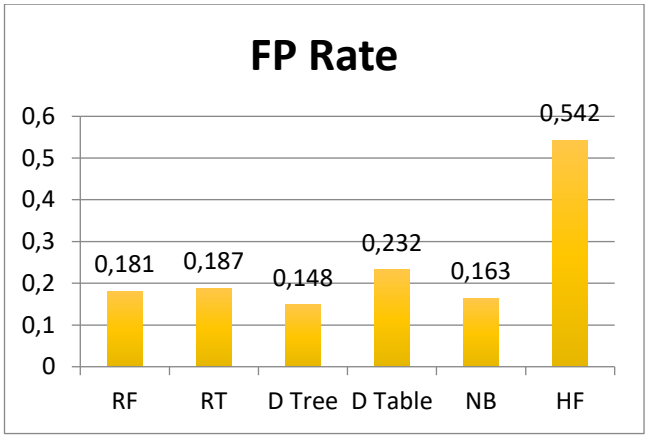


Fig.5. False positive rate of the applied techniques

Fig.5. shows the comparison of random forest, random tree, decision tree, decision table, naïve bayes and hoeffding tree in tems of false positive rate in which decision tree achieves lowest FPR of 0.148 among all other techniques.

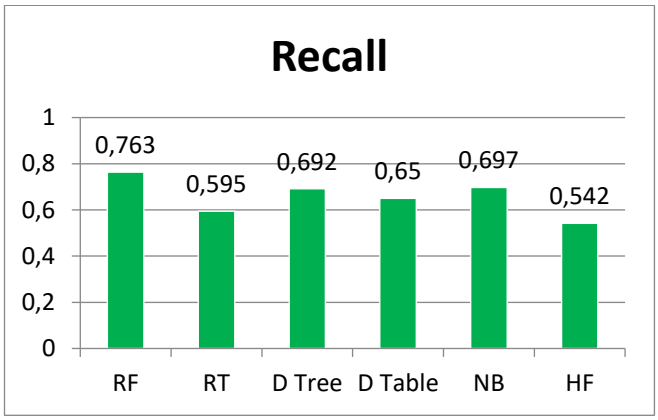


Fig.6. Recall of the applied techniques

Fig.6. shows the comparison of random forest, random tree, decision tree, decision table, naïve bayes and hoeffding tree in terms of recall in which RF achieves highest recall of 0.763 among all other techniques.

Fig.7. shows the comparison of random forest, random tree, decision tree, decision table, naïve bayes and hoeffding tree in terms of mean absolute error in which RF achieves lowest error of 0.0441 among all other techniques.

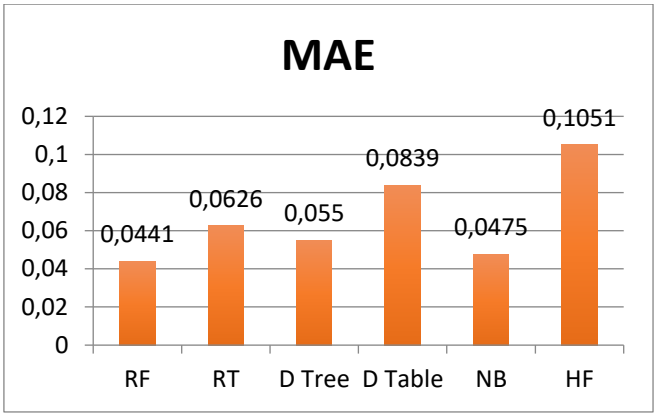


Fig.7. Mean absolute error (MAE) of the applied techniques

## VII. CONCLUSIONN

Unquestionably, one of the deadliest diseases that is difficult to accurately and precisely identify is chronic kidney disease. In summary, developing a tool for identifying chronic diseases will be helpful for both those who have trouble visiting a doctor and medical experts who need to solve urgent issues. Because CKD does not depend on a single trait, it is challenging to diagnose and difficult to predict. Additionally, the typical CKD symptoms do not significantly aid in diagnosing the condition. The system's apparent ability to solve invariant issues in the face of all odds and challenges is its most encouraging feature. The argument is made that even with a complicated dataset, huge numbers of distinguishable characteristics, and significant structural overlap, it still performs well in terms of accuracy. Given how much in the subject of machine learning remains unexplored, it is strongly advised that readers work on neural networks. As a result, this research article is only a first step, and any remaining questions can be answered over time. In summary, this work demonstrated the viability of using machine learning to assess the prognosis of CKD based on readily available data. When compared to other approaches, random forest showed comparable high predictability. Additionally, the sensitivity ratings of this ML model were higher, which might be helpful for patient screens. Future research will involve external validation and model enhancement with more predictor variables.

## REFERENCES

[1] R. Sawhney, A. Malik, S. Sharma, & V. Narayan, “A comparative assessment of artificial intelligence models used for early prediction and evaluation of

chronic kidney disease". Decision Analytics Journal, 6, 100169, 2023

[2] Hegde, Sandeepkumar, and M. R. Mundada "Early prediction of chronic disease using an efficient machine learning algorithm through adaptive probabilistic divergence based feature selection approach" International Journal of Pervasive Computing and Communications, 17(1), 20-36, 2021

[3] K. Xu, P. Rosenstiel, N. Paragas, C. Hinze, X. Gao, et al., "Unique Transcriptional Programs Identify Subtypes of AKI", Journal of the American Society of Nephrology 28(6):p 1729-1740, June 2017.

[4] A. Charleonnan, T. Fufaung, T. Niyomwong, W. Chokchueypattanakit, N. Ninchawee, et al., "Predictive analytics for chronic kidney disease using machine learning techniques". Management and Innovation Technology international conference (MITicon) (pp. MIT-80). IEEE, 2016

[5] V. Kunwa, C. Khushboo, A. Sai Sabitha, and A Bansal. "Chronic Kidney Disease analysis using data mining classification techniques." In 2016 6th International Conference-Cloud System and Big Data Engineering (Confluence), pp. 300-305. IEEE, 2016.

[6] Uma N Dulhare, Mohammad Ayesha. "Extraction of action rules for chronic kidney disease using Naïve bayes classifier", 2016, IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), 2016

[7] S. Elshahat, P. Cockwell, A. P. Maxwell, M.Griffin, T. O'Brien, et al., "The impact of chronic kidney disease on developed countries from a health economics perspective: A systematic scoping review." PLoS ONE 2020, 15, e0230512.

[8] J. R. Desai, S.V. Wang, and M. Vaduganathan, "Comparison of machine learning methods with traditional models for use of administrative claims with electronic medical records to predict heart failure outcomes." JAMA network open 3, no. 1 (2020): e1918962-e1918962.

[9] Chaurasia, Vikas, Mithilesh Kumar Pandey, and Saurabh Pal. "Chronic kidney disease: A prediction and comparison of ensemble and basic classifiers performance." Human-Intelligent Systems Integration 4, no. 1-2 (2022): 1-10.

[10] A. A. Asiri, B. Khan, F. Muhammad, S. U. Rahman, H. A. Alshamrani et al. "Machine Learning-Based Models for Magnetic Resonance Imaging (MRI)-Based Brain Tumor Classification", Computers, Materials & Continua, 36 (1), pp 1-14, 2023

[11] L. Bravo, N. Nieves-Pimiento, and K. Gonzalez-Guerrero, "Prediction of University-Level Academic Performance through Machine Learning Mechanisms and Supervised Methods," Ingeniería, vol. 28, p.e 19514, Nov. 2022, doi: 10.14483/23448393.19514

[12] F. Saberi-Movahed, M. Rostami, K. Berahmand, S. Karami, P. Tiwari, et al., "Dual regularized unsupervised feature selection based on matrix factorization and minimum redundancy with application in gene selection", Knowledge-Based Systems 256 (2022): 109884.

[13] J. Qin, L. Chen, Y. Liu, C. Liu, C. Feng, and B. Chen, "A machine learning methodology for diagnosing chronic kidney disease," IEEE Access, vol. 8, 2020, doi: 10.1109/ACCESS.2019.2963053.

[14] F. Muhammad, B. Khan, R. Naseem, A. A. Asiri, H. A. Alshamrani et al. "Liver Ailment Prediction Using Random Forest Model", Computers, Materials & Continua, 74 (1), 2023

[15] Theerthagiri, Prasannavenkatesan, and A. Usha Ruby. "RFFS: Recursive random forest feature selection based ensemble algorithm for chronic kidney disease prediction." Expert Systems 39, no. 9 (2022): e13048.

[16] S. Srivastava, R. K. Yadav, V. Narayan, and P. K. Mall. "An Ensemble Learning Approach For Chronic Kidney Disease Classification." Journal of Pharmaceutical Negative Results (2022): 2401-2409.

[17] H. Ilyas, S. Ali, M. Ponum, et al. "Chronic kidney disease diagnosis using decision tree algorithms", BMC Nephrol 22, 273 (2021).

[18] A. Sobrinho, A.C.M.D.S. Queiroz, L.L.D. SilvaCosta, M.E Pinheiro, A. Perkusich et al., "Computer-aided diagnosis of chronic kidney disease in developing countries: A comparative analysis of machine learning techniques", IEEE Access 2020, 8, 25407–25419.

[19] M. Hasan, S. Tahosin, A. Farjana, A. Sheakh, and M. Hasan. "A Harmful Disorder: Predictive and Comparative Analysis for fetal Anemia Disease by Using Different Machine Learning Approaches." In 2023 11th International Symposium on Digital Forensics and Security (ISDFS), pp. 1-6. IEEE, 2023.

[20] M. Z. Aljulifi, "Prevalence and reasons of increased type 2 diabetes in Gulf Cooperation Council Countries." Saudi Medical Journal 42, no. 5, 481-490, 2021

[21] A. Farjana, F. T. Liza, P. P. Pandit, M. C. Das, M. Hasan et al., "Predicting Chronic Kidney Disease Using Machine Learning Algorithms." In 2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC), pp. 1267-1271. IEEE, 2023.

[22] A. A. Asiri, F. Muhammad, H. A Alshamrani, K. Ullah, K. A Alshamrani et al., "Human Emotions Classification Using EEG via Audiovisual Stimuli and AI", Computers Materials & Continua, 73 (3), PP 1-15, 2022