# Comparative Analysis of Early Kidney Disease Detection Using Various Machine Learning Models

ASIF HASAN BIPLOB
*Computer Science*
*BRAC University*

K M ABRAR AHSAN
*Computer Science*
*BRAC University*

MD RUMMAN SHAHRIAR FAHIM
*Computer Science*
*BRAC University*

***Abstract*—The kidneys are the filtration system of the body. disease affecting the function of kidneys. Effective monitoring of CKD and efficient assessment of the risk of CKD are vital when it comes to protecting fragile patients from further damage to the kidney. Machine learning models will be able to help practitioners in accomplishing this goal efficiently due to its high speed in detecting and its high level of accuracy. Due to this, there is a massive influx of diagnosis systems and procedures in the healthcare circles that now employ machine learning given its capability of predicting diseases. We created and formulated illness predicting computer-aided designs for CKD diagnosis in this work. The CKD dataset was downloaded from UCL machine learning repository. it has several missing values that are filled up using the "mean/mode" and "random sampling method" techniques. 8 machine learning techniques (Random Forest, SVM, Naïve Bayes, Logistic Regression, KNN, XG_Boost, Decision Tree, and AdaBoost) were employed in the designing of models after the missing data was successfully retrieved. The techniques subsequently computed the performance evaluation comparisons between the result accuracies in order to establish the models for machine learning with maximum accuracies. Although KNN classifier model lands at the bottom with an accuracy of 73%, Random Forest and Logistic Regression have recorded impressive 99% accuracy, trailing Ada Boost, XG-Boost, Naive Bayes, Decision Tree, and SVM.**

***Keywords—Chronic kidney disease (CKD), Machine learning, Classification algorithms, Medical data mining, Healthcare***

## I. INTRODUCTION

Kidney also removes unwanted waste and extra fluid from our body. Apart from regulating correct blood concentration of salt, calcium and other important minerals and organic matters, the kidney also gets rid of acid produced by the cells of the body.

Their primary functions are removal of toxins from blood and making waste into urine. A kidney produces between 1-1.5 liters of urine and about 160 grams of weight per day [1]. is a vital organ of our body and over the course of a day, the two kidneys collaborate to extract 200 liters of fluid from the blood. But occasionally, a sickness may assault it. If the kidneys are unable to eliminate waste that has been tainted by poisons, renal failure may be fatal. The kidneys' functions include filtering blood through the bladder and eliminating toxins from the body.

Urination. There are two categories of kidney disorder. acute and chronic [2]. One of the greatest issues of the world economy is the disease of kidneys. This disease becomes two to three percent of the yearly income of developed nations. Chronic kidney diseases are conditions that affect the kidneys and impair the ability of them to take care of our health. Waste can accumulate to the maximum levels in our blood, if a kidney disease deteriorates, which can lead to such problems as Nerve damage, weak bones, anemia, Hypertension and poor nutrition. The risk of vascular and heart diseases is also increased in cases of kidney dysfunction.

CKD is a highly morbid and mortal condition affecting people everywhere in the world [3]. Therefore, the early diagnosis and identification of chronic kidney disease (CKD) are all important. It may allow people to receive timely care to avoid the spread of the illness. In controlling their meals the chronic kidney disease (CKD) patients also have to consider other factors including blood potassium, urea, calcium, phosphorus among others. Because of the availability of biological data, the machine learning methods are used more and more often in the healthcare industry in order to create disease prediction models. Moreover, such techniques as deep learning and ensemble learning enhanced the predictive ability of machine learning models significantly. Extracted information from electronic health record (EHR) can be harnessed to develop accurate illness prediction models [4][5]. Thus, ML technologies can serve as an indispensable device for predicting various kidney ailments and handy tool for fulfilling medical attention to the people who need it the most if there is enough data existing. However, from a public health paradigm, the lack of universal availability to lab data is hampering the computer-assisted medical technology development.

Today, one can find many studies on all kinds of heart, liver, and kidney diseases and identification and classification of various diseases such as cancer among others. More precisely, in 2004 machine learning was believed to be the most influential means of detecting the variety of cancer-associated issues [6]. With increasing use of ML applications in disease prediction, it is apparent how much the healthcare epidemiology is experiencing change. In an attempt to forecast breast cancer, Huang et al. [7] investigated two datasets and measured the capabilities of individual support vector machine (SVM) classifiers and SVM classifier ensembles. They used a variety of various kernel functions in their study as they are able to have a significant effect on the efficacy of the SVM classifier. The results indicated that linear kernel-based SVM ensemble is superior to the ones constructed on the RBF kernel in the case of small and large datasets. Another impressive work

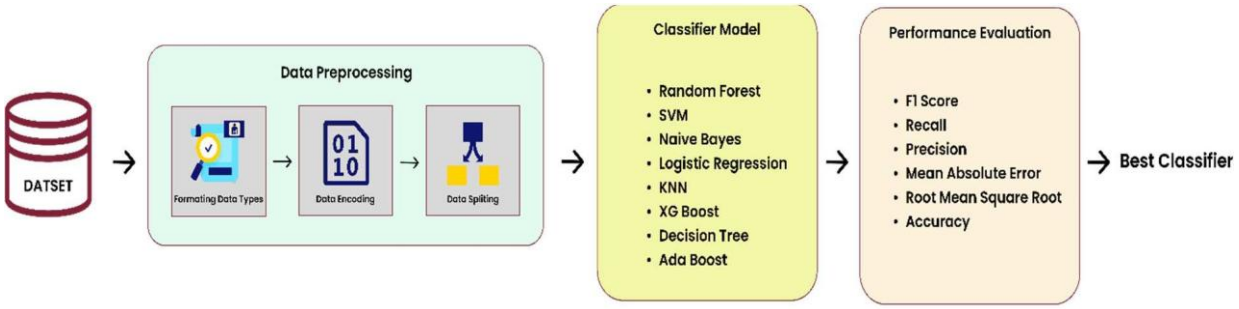


Fig. 1. Workflow of Chronic Kidney Disease Prediction Task

is the use of the ML approach by Shah et al. [8] to predict cardiovascular disease with the help of multiple supervised algorithms. The purpose of the present study was to identify the most effective machine learning method that would be the best cardiovascular disease predictor. KNN performed better than the other algorithms in the proposed model and achieved 90.78% accuracy in the training data.

As a short form, CKD means chronic kidney disease, which has been one of the major causes of deaths in the recent years. Worldwide, the death caused by CKD rose by 41.5% during 1990-2017 [9]. Consequently, it could be more rational to decide on the reason behind it or the sickness. Random Forest, with 99.75% diagnostic accuracy, seemed to be the best machine learning model out of six possessed by Qin et al. [10] to diagnose kidney illness; Nevertheless, due to the errors generated by the proposed model, researchers generated a new model involving the random forest and logistic regression by means of a perceptron. Accuracy of 99.83% was achieved by this model. Three different models are used in the tests conducted by Chen et al. [11] to determine the best methods in the identification of CKD. SIMCA was a little bit below in their first trial than KNN and SVM which had higher accuracy that was almost 100%. They also ran further testing on the same three algorithms and composed data with randomly added noise and the results proved that SVM was 99 percent accurate.

In 2016, Charleonnan et al. [12] have applied SVM methods to develop a CKD identification. They employed a wide range of machine learning classifiers such as SVM, KNN, logistic regression, and decision trees in their experiments to experiment with the best accurate way of recognizing CKD. For each of the above models, the respective results of classification were 98.3%, 98.1%, 96.6, and 94.8%. Shamrat et al. [13] studied the effectiveness of supervised machine learning on the pre-processed datasets.

Techniques to find the best performing renal disease classifier: Random Forest, Decision Tree, KNN, Logistic Regression. In the report, the study revealed that Random Forest and Logistic Regression both had an accuracy rate of 100 %. Imran et al. suggested a computer-assisted prediction of a disease for kidney diagnostics [14]. Three separable machine learning (ML) methods – feedforward neural networks, logistic regression, as well as wide and deep learning, were run on two different data sets. The results of the experiments showed that balanced and unbalanced samples could be processed by feedforward neural networks rather fulfilling.

It is our aim to give an idea on how CAD can be used to diagnose chronic renal disease and make better decisions in the healthcare environment. Also, we would want to demonstrate that in an attempt to establish which classifier is most suited to our model we used assorted supervised machine learning algorithms and measured their effectiveness. Below are organization of the variety of parts of the paper. Section II demonstrates an integrated approach with the dataset used, the data pretreatment steps, short description of the advised methodology, and the explanation of other machine learning methodologies. Section III gives examples of the results and discussion, evaluation metrics, and results comparison. Finally, after closing statement, section IV ends.

## II. Materials and Methods

Theoretical validity and the clarity of the research are the matters discussed in this part. For the study, the CKD data from UCI machine learning repository is used. The data was cleaned with the help of different preprocessing techniques. After the selecting the features of the dataset, the data were cleaned, processed and split into training and test sets. 8 machine learning algorithms were trained with training data. The algorithms were used to forecast test data after it had been trained. Fig. 1 illustrates the whole procedure of the assessment of the model for prediction of CKD.

### *A. Data Acquisition*

This dataset is taken from one of the famous machine learning resources called UCI [15]. For this dataset, there are 400 unique data points all in all. In this dataset, there are 25 different traits that comprise of 250 traits labeled as “CKD” while 150 traits are labeled as “not CKD”. In this dataset, there is the existence of two separate kinds of data. object (14) and float64 (11). In the dataset, there are two parts: 80% of the dataset is used for training and 20% for testing purposes. The distribution of the dataset as given earlier is demonstrated in Fig. 2.

### *B. Data Preprocessing*

Working with raw and noisy data is impractical, necessitating preprocessing through "Feature Encoding" and "Formatting Data Types." For handling missing values, the study employed "Random Imputation Method" and "Mean/Mode Sampling method." Several columns required conversion to appropriate data types. The first data frame contains numerical values, while the second comprises categorical data, constituting feature encoding. Data was renamed for improved manageability. In the "classification" column, "CKD" was transformed to "0" and "not CKD" to "1," assigning numerical values to entries, which facilitates more efficient computational processing. Missing information was addressed via two approaches: mean-mode sampling for minimal null values and random sampling for substantial null values [16]. Following the elimination of columns with null

values, the dataset contained both categorical and numerical columns, with categorical data converted to numeric values through feature encoding.

Here Fig. 3 illustrates the distribution of numerical features in the dataset.

| Features | Number of missing Values | Data Type |
|---|---|---|
| age | 9 | float64 |
| blood_pressure | 12 | float64 |
| specific_gravity | 47 | float64 |
| albumin | 46 | float64 |
| sugar | 49 | float64 |
| red_blood_cells | 152 | object |
| pus_cell | 65 | object |
| pus_cell_clumps | 4 | object |
| bacteria | 4 | object |
| blood_glucose_random | 44 | float64 |
| blood_urea | 19 | float64 |
| serum_creatinine | 17 | float64 |
| sodium | 87 | float64 |
| potassium | 88 | float64 |
| haemoglobin | 52 | float64 |
| packed_cell_volume | 70 | object |
| white_blood_cell_count | 105 | object |
| red_blood_cell_count | 130 | object |
| hypertension | 2 | object |
| diabetes_mellitus | 2 | object |
| coronary_artery_disease | 2 | object |
| appetite | 1 | object |
| peda_edema | 1 | object |
| aanemia | 1 | object |
| class | 0 | object |

Fig. 2. Chronic kidney disease (CKD) dataset insights overview

### *C. Data Splitting*

After data processing, the dataset needed to be divided into two parts, which is; industries. test and train. The first performance test we use is associated with training our models on the training dataset followed by testing them on the same. What is used in testing will be twenty percent of the dataset while training will take eighty percent of the dataset.

### *D. Selecting Models and Evaluation*

Now, that we have divided the dataset, we are going to choose our machine learning model and start training. Some of the models that were employed included; Random Forest, Support Vector Machine (SVM), Naive Bayes, Logistic regression, K-Nearest Neighbor, XGBoost, decision tree, and AdaBoost. We evaluate the whole classification report of each machine learning model during the evaluation phase. In addition, it was also estimated the MAE and the RMSE. For further machine learning models' analysis, confusion matrix graphs and ROC curves have been drawn up.

### *E. Machine Learning Classifiers*

#### *1) Decision Tree*

Both discrete and continuous properties can be predicted using the Decision Tree technique [17]. It is a method of regression and classification. The method uses the relationships between the columns of a dataset to predict discrete qualities. Based on their values, or states, it forecasts the states of the columns you provide are predictable.

The input columns that are connected to the predicted column are identified by the procedure. Because the decision tree resembles the stages someone would follow while making a real-life decision, it is simple to understand. When it comes to decision-making, it could be really beneficial. Considering every potential solution to an issue is a smart approach [18]. Cleaning the data is less crucial with this approach than it is with others.

#### *2) K-Nearest Neighbor*

The KNN (K-nearest neighbors) is an easy supervised method [19]. It can be employed in regression and classification issues. However, it is most often put to use on problems of categorization. Because K-Nearest Neighbors makes use of all available data for training, without a distinct stage of training, it involves a nonparametric pattern of learning. In addition, it overlooks the underlying data. As KNN does not have a model, it does not learn and keeps the whole dataset. Because it compares K's neighbors every time that fresh data is provided to predict the outcome, the value of K is important. The measurement of the distance between two previously assorted data sets is determined. From the distance, it is possible to find the most similar data set that is similar to the new one. There is a use of Euclidean geometry while arriving at the distance.

#### *3) Logistic Regression*

One of the most common statistical estimation techniques in modeling of two alternative outcomes is logistic regression [20]. The logistic regression in statistical research is done through various learning techniques.

The LR algorithm was developed based on a particular application of the neural network technology. This approach, although easier to set up and run, has a lot in common with the neural networks. It is possible to forecast dependent variable with categorical value by use of logistic regression. So, the output from the function must be a list of categories or a list of single values. The chances are given by numbers from 0 to 1, but they may be true or false, 0 or 1, etc.

#### *4) Naïve Bayes*

The Naïve Bayes methodology implements Bayes theorem in classifier training [21]. Essentially, this approach establishes foundations for educating a probabilistic classifier. It generates a probability distribution across multiple classes for any given observation.

#### *5) Support Vector Machine*

SVMs represent one of the most prevalent supervised learning methodologies. These facilitate the resolution of regression and classification challenges [22]. However, their principal utilization occurs within machine learning for addressing classification problems. The SVM approach identifies the optimal line or decision boundary that divides the n-dimensional space into classes, thereby facilitating straightforward classification of subsequent data points.

#### *6) Random Forest*

Random Forest (RF) functions as a widely implemented algorithm in supervised learning applications. Within machine learning contexts, it addresses both regression and classification issues [23]. This technique builds upon ensemble learning, a strategy that integrates numerous classifiers to tackle complex problems and enhance model performance. As a classifier, it employs multiple decision trees on various dataset subsets and aggregates their results, aiming to improve the dataset's anticipated accuracy. Rather than depending on an individual decision tree for determinations, Random Forest utilizes majority voting from tree predictions to draw conclusions.

*7) XGBoost*

XGBoost algorithm constructs decision trees in a consecutive manner [24]. With XGBoost, weights are important. Independent variable before being converted into forecasting decision tree is given a weight. Variables that were misjudged by the previous tree are more heavily weighted before being fed into the second tree. These various classifiers and predictors are used to generate a reliable and strong model by combining them. It can be used to solve regression, classification, ranking and custom prediction types of problems among others.

*8) AdaBoost*

Boosting methods are methods to achieve strong classifiers by combining other weak classifiers in order to enhance classification accuracy [25]. Another helpful technique called: Adaptive Boosting, was introduced by Friedman and Schicher in 1997, but it was only in 2000 when Jacob Schapire suggested using it for classification. It was shown that in order to deal with this situation other Boost applied more generalizations. Boosting algorithms cover various medical problems, such as the recognition of the protein structure class, cancer detection, and breast cancer diagnosis.

Absolute Error (MAE), F1-Score, Root Mean Square Error (RMSE), Precision And, Recall. We are going to discuss the evaluation measures and the outcome in this part.

It is possible to see the True Positive, True Negative, False Positive, and False Negative values using the examinations of the confusion matrices [26][27], which are used to calculate the metrics. Cases, where the values are positive, express the fact that a person has CKD and the negative values state that a person does not have CKD.

- ***TP*** – Refers A set of CKD samples that were correctly predicted.
- ***FP*** – Refers to some anticipated samples that are not CKD but were believed to be.
- ***TN*** – Refers to several precisely anticipated non-CKD samples.
- ***FN*** – Refers to several CKD-containing samples that were mislabeled as non-CKD.

This suggested CKD classifiers are identified by examining the prediction ratio. An overview of each model's performance is given in Table I and Fig. 5. Additionally, Figs. 4 and 6 display the confusion matrix and ROC curve performance. Table I represents the sum of the performance of each of the supervised

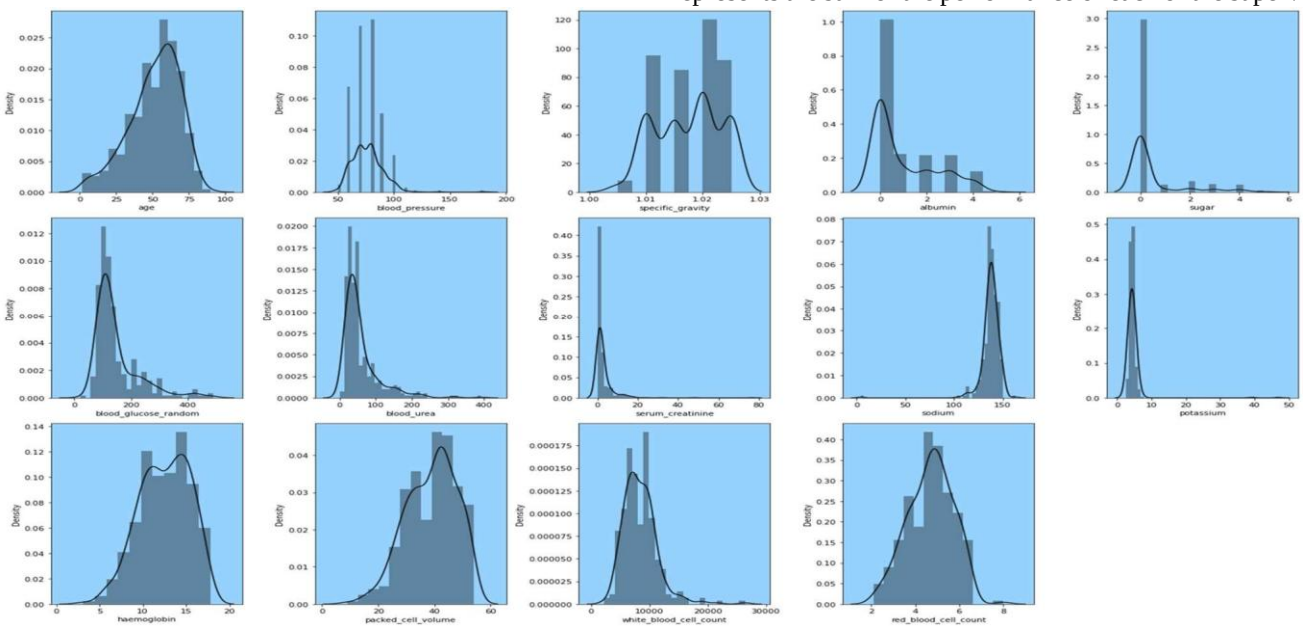


Fig. 3. Numerical Feature Distribution of Raw CKD dataset

## III. Results and Discussion

To our help in this experiment, we utilize the data set from the famous UCI machine learning repository. From 400 data with 25 distinct attributes, there are 250 data which were labeled as "CKD" and 150 data with "not CKD". After then the split of the dataset was in the ratio 80:20 between the testing and training sets. After sufficient research we have chosen a variety of machine learning models such as the SVM, Naive Bayes, Random Forest, Decision tree, Ada Boost, Logistic Regression and XG Boost so as to accomplish our task.

### *A. Models Performance Evaluation and Comparison*

This study aims at measuring various machine learning classifiers to determine Chronic Kidney Disease (CKD). In this manner, the above-mentioned models' performance is evaluated through five different evaluation measures. Mean

machine learning method after various trials. It is obvious from the performance matrices that most of these models –except KNN – did a damn good job on dataset presented by us.

Random Forest was the best performers with 0.83 as the highest level of classification as shown on the performance table. ADA – Boost, XG – Boost, Decision Tree and SVM all had significant percentages of classification attainments with accuracy as high as 0.37, 0.37, and 0.36 KNN, nonetheless, performed also well with F1-Measures and classification accuracy (0.64) as compared to the other models.

The confusion matrices depicting classification performance of proposed classifiers are shown in Fig. 2, while Fig. 1 presents an overall performance comparison bar chart. Examination of these matrices reveals Random Forest and Logistic Regression achieve nearly accurate TP and TN predictions.

TABLE I. PERFORMANCE MEASURE OF ML CLASSIFIERS

| Classifier Models | Class Type | Precision | Recall | F1-Score | MAE | RMSE | Accuracy |
|---|---|---|---|---|---|---|---|
| Random Forest | CKD | 1.00 | 0.74 | 0.85 | 0.033 | 0.821 | 0.83 |
| | Not CKD | 0.68 | 1.00 | 0.81 | | | |
| SVM | CKD | 0.64 | 1.00 | 0.78 | 0.072 | 0.224 | 0.64 |
| | Not CKD | 0.90 | 0.96 | 0.93 | | | |
| Naïve Bayes | CKD | 0.64 | 0.98 | 0.78 | 0.025 | 0.158 | 0.64 |
| | Not CKD | 0.96 | 0.96 | 0.96 | | | |
| Logistic Regression | CKD | 0.65 | 0.99 | 0.79 | 0.0125 | 0.112 | 0.66 |
| | Not CKD | 0.80 | 0.70 | 0.13 | | | |
| KNN | CKD | 0.64 | 1.00 | 0.78 | 0.275 | 0.524 | 0.64 |
| | Not CKD | 0.57 | 0.69 | 0.69 | | | |
| XGBoost | CKD | 0.32 | 0.50 | 0.39 | 0.025 | 0.158 | 0.37 |
| | Not CKD | 0.36 | 1.00 | 0.53 | | | |
| Decision Tree | CKD | 1.00 | 0.10 | 0.20 | 0.037 | 0.194 | 0.36 |
| | Not CKD | 0.68 | 1.00 | 0.53 | | | |
| AdaBoost | CKD | 1.00 | 0.10 | 0.20 | 0.025 | 0.158 | 0.37 |
| | Not CKD | 0.36 | 1.00 | 0.53 | | | |

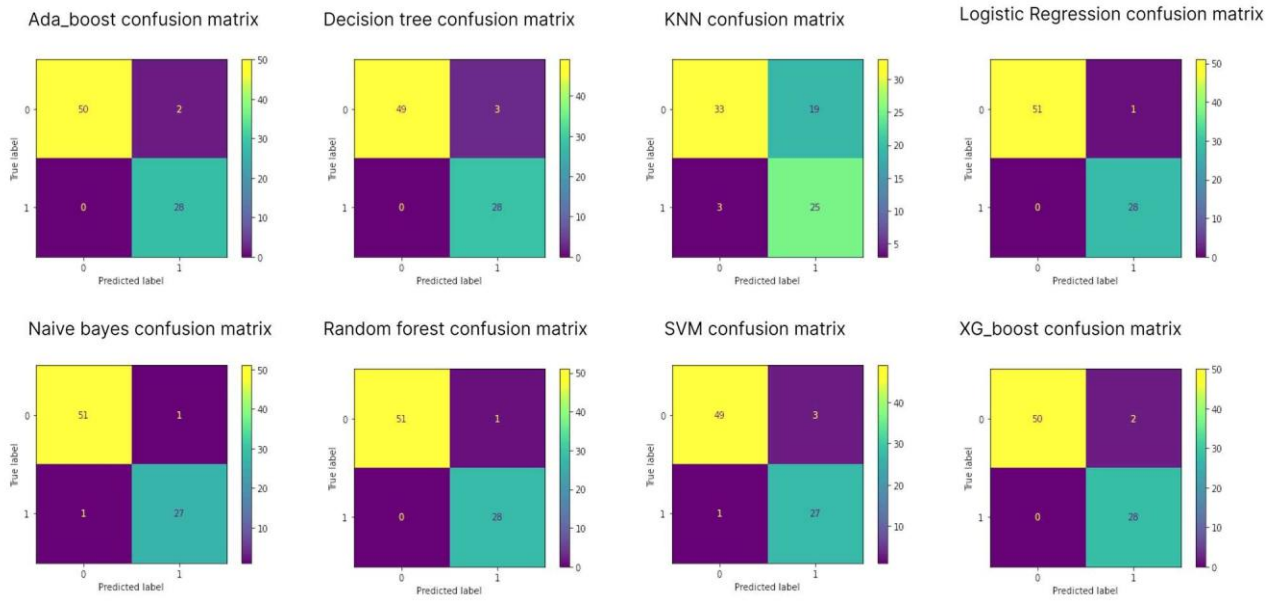


Fig. 4. Confusion matrices of ML classifiers in CKD prediction

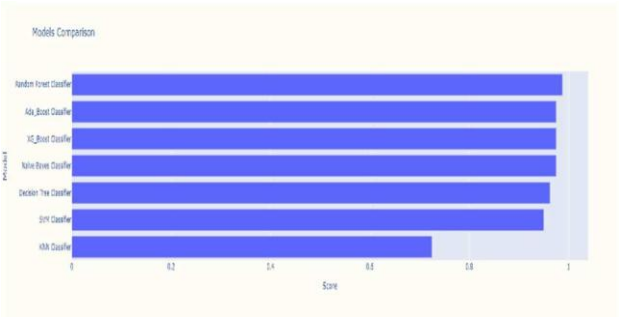


Fig. 5. Overall classification accuracies of the ML classifiers

### *B. Performance Comparison with the Existing Models*

Comparing to the results we got from this paper, other models that trained in the same genre have been compared. From the table, it is possible to find a common classifier, that is, the Random Forest. It has a high degree of accuracy on several projects with different sets of data. In this case, we compared the results of our suggested work with various dimensions too instead of just comparing our work with the same parameter, i.e., CKD. In an Aushtmi et al study on heart disease prediction, Random Forest performed better as compared to other techniques [28]. Thus, if speaking about the accuracy and efficiency of the diagnosis of CKD, we can say that Random Forest classifier is the best one of all of them.

## IV. CONCLUSION

The aim of our research was to employ eight different machine learning algorithms namely Random Forest, SVM, Naive Bayes, Logistic Regression, KNN, XG Boost, Decision Tree, and AdaBoost, whose task was to predict correctly for CKD. Five assessment metrics namely Recall, Precision, Mean Absolute Error, Root Mean Square Error and the best-performing machine learning algorithms were Random Forest and Logistic Regression. On the dataset presented, these two models delivered 99 percent accuracy, with the last place being taken by KNN. KNN was not accurate and performed poorly with an accuracy of 73%. Our project's classifier method entails an accuracy rate adequate to our specifications. It still needs to be trained very well and tested using various datasets before it can be used for production purposes. We expect that as being the successful outcome of the current study, the subsequent disease prediction by means of machine learning will find its way, thus to contribute to developing health care systems.

## REFERENCES


[1] C. Barbieri *et al.*, "A new machine learning approach for predicting the response to anemia treatment in a large cohort of End Stage Renal Disease patients undergoing dialysis," *Comput. Biol. Med.*, vol. 61, pp. 56–61, 2015.

[2] N. R. Hill *et al.*, "Global prevalence of chronic kidney disease--a systematic review and meta-analysis," *PLoS One*, vol. 11, no. 7, p. e0158765, 2016.

[3] M. Alloghani, D. Al-Jumeily, T. Baker, A. Hussain, J. Mustafina, and A. J. Aljaaf, "Applications of machine learning techniques for software engineering learning and early prediction of students' performance," in *International conference on soft computing in data science*, 2018, pp. 246–258.

[4] A. Callahan and N. H. Shah, "Machine learning in healthcare," in *Key Advances in Clinical Informatics*, Elsevier, 2017, pp. 279–291.

[5] A. J. Collins *et al.*, "Who should be targeted for CKD screening? Impact of diabetes, hypertension, and cardiovascular disease," *Am. J. Kidney Dis.*, vol. 53, no. 3, pp. S71--S77, 2009.

[6] J. F. Mccarthy *et al.*, "Applications of machine learning and highdimensional visualization in cancer detection, diagnosis, and management," *Ann. N. Y. Acad. Sci.*, vol. 1020, no. 1, pp. 239–262, 2004.

[7] M.-W. Huang, C.-W. Chen, W.-C. Lin, S.-W. Ke, and C.-F. Tsai, "SVM and SVM ensembles in breast cancer prediction," *PLoS One*, vol. 12, no. 1, p. e0161501, 2017.

[8] D. Shah, S. Patel, and S. K. Bharti, "Heart disease prediction using machine learning techniques," *SN Comput. Sci.*, vol. 1, no. 6, pp. 1–6, 2020.

[9] B. Bikbov *et al.*, "Global, regional, and national burden of chronic kidney disease, 1990--2017: a systematic analysis for the Global Burden of Disease Study 2017," *Lancet*, vol. 395, no. 10225, pp. 709–733, 2020.

[10] 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, pp. 20991–21002, 2019.

[11] Z. Chen, X. Zhang, and Z. Zhang, "Clinical risk assessment of patients with chronic kidney disease by using clinical data and multivariate models," *Int. Urol. Nephrol.*, vol. 48, no. 12, pp.

[12] F. M. J. M. Shamrat, P. Ghosh, M. H. Sadek, M. A. Kazi, and S. Shultana, "Implementation of machine learning algorithms to detect the prognosis rate of kidney disease," in *2020 IEEE International Conference for Innovation in Technology (INOCON)*, 2020, pp. 1–7.

[13] A. Al Imran, M. N. Amin, and F. T. Johora, "Classification of chronic kidney disease using logistic regression, feedforward neural network and wide \& deep learning," in *2018 International Conference on Innovation in Engineering and Technology (ICIET)*, 2018, pp. 1–6.

[14] A. Frank, A. Asuncion, and others, "UCI machine learning repository, 2010," *URL http//archive. ics. uci. edu/ml*, vol. 15, p. 22, 2011.

[15] E. M. Hernández-Pereira, D. Álvarez-Estévez, and V. Moret-Bonillo, "Automatic classification of respiratory patterns involving missing data imputation techniques," *Biosyst. Eng.*, vol. 138, pp. 65–76, 2015.

[16] H. Cho *et al.*, "A similarity study of content-based image retrieval system for breast cancer using decision tree," *Med. Phys.*, vol. 40, no. 1, p. 12901, 2013.

[17] S. Sakib, A. K. Tanzeem, I. K. Tasawar, F. Shorna, M. A. B. Siddique, and S. B. Alam, "Blood Cancer Recognition Based on Discriminant Gene Expressions: A Comparative Analysis of Optimized Machine Learning Algorithms," in *2021 IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, 2021, pp. 385–391.

[18] M. F. Faruque, I. H. Sarker, and others, "Performance analysis of machine learning techniques to predict diabetes mellitus," in *2019*.

[19] R. Xi, N. Lin, and Y. Chen, "Compression and aggregation for logistic regression analysis in data cubes," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 4, pp. 479–492, 2008.

[20] U. N. Dulhare and M. Ayesha, "Extraction of action rules for chronic kidney disease using Na{\"\i}ve bayes classifier," in *2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, 2016, pp. 1–5.

[21] S. Sakib, M. A. B. Siddique, and M. A. Rahman, "Performance Evaluation of t-SNE and MDS Dimensionality Reduction Techniques with KNN, ENN and SVM Classifiers," in *2020 IEEE Region 10 Symposium (TENSYMP)*, 2020, pp. 5–8.

[22] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[23] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.

[24] R. E. Schapire, "Explaining adaboost," in *Empirical inference*, Springer, 2013, pp. 37–52.

[25] S. Sakib, N. Yasmin, A. K. Tanzeem, F. Shorna, S. B. Alam, and others, "Breast Cancer Detection and Classification: A Comparative Analysis Using Machine Learning Algorithms," in *Proceedings of Third International Conference on Communication, Computing and Electronics Systems*, 2022, pp. 703–717.

[26] T. Anjum, T. E. Chowdhury, S. Sakib, and S. Kibria, “Performance Analysis of Convolutional Neural Network Architectures for the Identification of COVID-19 from Chest X-ray Images,” in *2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)*, 2022, pp. 446–452.

[27] A. Deb, M. S. A. Koli, S. B. Akter, and A. A. Chowdhury, “An Outcome Based Analysis on Heart Disease Prediction using Machine Learning Algorithms and Data Mining Approaches,” in *2022 IEEE World AI IoT Congress (AIIoT)*, 2022, pp. 1–7.

[28] M. Wickramasinghe, D. M. Perera, and K. Kahandawaarachchi, “Dietary prediction for patients with chronic kidney disease (CKD) by considering blood potassium level using machine learning algorithms,” in *2017 IEEE Life Sciences Conference (LSC)*, 2017, pp. 300–303.

[29] A. Saha, A. Saha, and T. Mittra, “Performance measurements of machine learning approaches for prediction and diagnosis of chronic kidney disease (CKD),” in *Proceedings of the 2019 7th international conference on computer and communications management*, 2019, pp. 200–204.