JCBI
Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
Volume 06 Issue 02
2024

Research Article
https://doi.org/10.56979/602/2024

Predictive Modeling of Chronic Kidney Disease Using Extra Tree Classifier: A Comparative Analysis with Traditional Methods

Muhammad Imran[1], Naeem Aslam[1], Haroon Ahmad[2*], Faheem Mazhar[1], Yousuf Iqbal Bhatti[1], and Umair Abid[1]

[1]Department of Computer Science, NFC-IET, Multan, Pakistan.
[2]Department of Computer Science, Air University Islamabad, Multan Campus, Multan 60000, Pakistan.
*Corresponding Author: Haroon Ahmad. Email: haroon@aumc.edu.pk

Received: January 05, 2024 **Accepted:** February 26, 2024 **Published:** March 01, 2024

Abstract: With a high prevalence of morbidity and death, chronic renal illness is a major global health concern. Conventional diagnostic methods frequently miss the disease until it has grown to an advanced stage, despite the fact that prompt diagnosis and treatment can greatly improve patient outcomes. This work suggests a unique method for utilizing machine learning (ML) algorithms to identify kidney sickness, which might offer a solution to this urgent healthcare problem. One of the many industries where machine learning—a subset of artificial intelligence—has demonstrated great potential is healthcare. Due of its capacity to make predictions and Take note of the data, it is a useful tool for predicting illnesses. This study uses a variety of clinical indicators along with machine learning methods to predict when chronic kidney disease (CKD) will manifest. The proposed model uses a dataset comprising numerous patient records with various attributes. These attributes are used as input features for the machine learning algorithm. The target variable is the presence or absence of chronic kidney disease. A number of machine learning algorithms are used, and their performances are contrasted, including KNN, Chronic Kidney Disease, Machine Learning, Gradient Boosting Classifier, Ada Boost Classifier, Random Forest Classifier, XgBoost, Cat Boost, and Extra Trees Classifier. To assess each algorithm's predictive accuracy, sensitivity, specificity, and other performance metrics, a subset of the dataset is used for training, and afterwards the algorithm is tested using untested data. The findings show that machine learning algorithms, some of which are more accurate than others, can predict chronic kidney disease. According to these results, machine learning may prove to be a useful tool in the early diagnosis of chronic kidney disease, allowing for prompt intervention and maybe leading to better patient outcomes.

Keywords: Chronic Kidney Disease; Machine Learning; KNN; Decision Tree; Random Forest; XgBoost; Extra Tree.

1. Introduction

The prevalence of chronic kidney disease, a non-communicable disease, has sharply increased patient admission, morbidity, and death rates worldwide [1]. It is becoming more widespread quickly and becoming well-known as among the main causes of death worldwide. A survey conducted between 1990 and 2013 found that chronic renal disease has increased the annual death rate, making it the leading cause of morality globally. [2]. A number of variables indicate that 850 million people worldwide are anticipated to be impacted by kidney ailments. According to a 2019 WKD study, at least 2.4 million individuals die each year from CKDrelated disorders. Currently, it ranks as the sixth fastest-rising cause of death around

the globe [3]. Persistent renal illness is a challenging health policy concern as it is growing more commonplace worldwide. Because there is still little awareness, prevention, and treatment, the burden is significantly greater in low-income nations [4]. In Ethiopia, A serious public health concern that impacts hundreds of thousands of people of all ages and genders is kidney disease. [5]

It is thought that a lack of healthy diet, physical activity, and safe water have all contributed. Furthermore, the general public in rural areas knows very little about chronic kidney disease. 4,875, according to a 2017 WHO report, were Ethiopian deaths attributable to kidney illness. With 0.77% of all deaths, the nation is ranked 138th in the world. The country was rated 109th in 2018 due to a rise in the death rate adjusted for age, which was 8.46 per 100,000 population to 12.70 per 100,000 [6]. The lowered Glomerular Filtration Rate (GFR), which gauges kidney function, and aberrant renal function are the basis for the National Kidney Foundation's classification of chronic kidney disease stages into five. Stages 1 and 2 are the least severe, exhibiting very few symptoms, On the other hand, stage 5 renal failure is considered to be end-stage. Renal Replacement Therapy (RRT) is expensive for treating total renal failure. The majority of impoverished nations, including Ethiopia, do not have access to this medication. Because of the lack of facilities, doctors, and affordable treatment options, renal failure and associated consequences are difficult to manage in poor nations [7].

Thus, early detection of chronic renal disease is essential for minimizing financial burden and maximizing the effectiveness of treatment [8]. By using machine learning techniques to predictive analysis, early detection of chronic kidney disease can enable efficient and timely therapy. Decision-Tree, Random-Forest, and Support-Vector-Machine were the methods utilized in this study to diagnose chronic renal disease. The majority of earlier studies concentrated on just two classes, which makes it challenging to propose a course of treatment because it depends on the severity of chronic kidney disease [9]. Chronic kidney disease can be influenced by conditions such as diabetes, hypertension, and heart disease. Kidney disease may also be influenced by a family history of renal failure. Loss of appetite, vomiting, and weight loss are signs of kidney illness. Early kidney disease prediction can aid in rectification, although this is not achievable. preventing significant harm. We need information on a few variables that have a strong correlation with renal illness in order to forecast this. By analyzing the data on those indices, applying five machine learning classification approaches, and choosing the one that has the highest accuracy rate for disease prediction, our goal is to make predictions about renal disease [10].

2. Problem Statement

To develop an effective machine learning model for early detection and prediction of chronic kidney disease using various clinical indicators and patient data. Chronic kidney disease has high morbidity and mortality rates worldwide. Traditional diagnostic techniques often fail to detect the disease until it has progressed to an advanced stage. Early detection and prompt intervention can significantly improve patient outcomes for persistent renal illness. The study proposes using algorithms for machine learning trained on patient data with attributes like age, blood pressure, specific gravity, albumin, etc. to predict the presence or absence of chronic kidney disease. Addressing this problem of early detection through machine learning techniques could prove to be a useful tool and potentially lead to better management of chronic kidney disease. So, in essence, the problem being addressed is the timely and reliable early detection and prognosis of chronic renal disease leveraging machine learning models and patient data, to enable prompt treatment and improve outcomes.

3. Literature review

Many machine learning algorithms have been effectively used to the classification of chronic renal illness using patient data. A Chronic Kidney Illness dataset from India was utilized by Charleonnan et al. [11] to evaluate a number of predictive models, including logistic regression, K-nearest-neighbors, support vector machines, and decision trees. To determine which classifier is most effective in forecasting chronic kidney disease. They found that the most accurate categorization was SVM. (98.3%) and the maximum sensitivity. Salekin and Stankovic [12] assessed classifiers such as KNN, RF(Random Forest), and ANN using a CKdataset of 400 instances. A choice of wrapping features approach was utilised to pick five characteristics for the study's model construction. 98% classification accuracy by RF with an RMSE of 0.11

is the highest. S. Tekale et al. [13] used a dataset with 450 occurrences and 14 features to work on the project "Prediction of Chronic Kidney Disease Using Machine Learning Algorithm." They have employed support vector machines (SVM) and decision trees(dt). After preprocessing the dataset, there are now just 14 characteristics instead of 25. SVM is recognized as a superior model, with a 96.75% accuracy rate.

Xiao et al. [14] suggested utilizing neural networks, XGBoost, random forests, logistic regression (LR), Elastic Net (EN), lasso regression (L Reg), ridge regression, support vector machines (SVM), and k-nearest neighbor to predict the progression of chronic kidney disease. The models were then compared depending on how well they performed. They categorized the result as mild, moderate, or severe using the medical histories of 551 patients who had proteinuria and 18 characteristics. They have determined that, sensitivity and specificity were 0.83 and 0.82, respectively, while the AUC was 0.873., logistic regression outperformed other methods. Using machine learning, Mohammed and Beshah [15] investigated the creation of a self-learning knowledge-based system for the diagnosis and management of the initial three stages of chronic renal disease. A prototype that allows patients to ask KBS to observe the advice transmission was constructed using a limited set of data for this study. They generated the rules using a decision tree. The prototype's overall accuracy has been reported to be 91%.

Priyanka et al. [16] used naïve bayes to predict chronic renal disease. After testing with various Naïve Bayes had a greater accuracy of 94.6% than the other algorithms, including KNN, SVM, ANN, and decision trees. In their study, Almasoud and Ward [17] sought to determine whether a subset of features might be used to assess the predictive power of machine learning algorithms for chronic renal disease. To choose predictive features, they employed Cramer's V test, ANOVA, and Pearson correlation [18]. They have used machine learning methods including LR, SVM, RF, and GB for modeling. Ultimately, they came to the conclusion that, with an F-measure of 99.1%, Gradient Boosting offers the best accuracy. By looking at CKD patient data, Yashfi [19] suggested applying machine learning techniques to predict CKD risk. AN-N and RF have been used. Twenty of the twenty-five features were retrieved, and ANN and RF were used. With an accuracy of 97.12%, RF has been chosen as the most accurate [20].

According to reviews above, a number of studies [21-22] on machine learning techniques for chronic renal disease prediction have been carried out. A number of factors, such as the quantity, quality, and duration of the dataset collection, are crucial in enhancing model performance. Using large-scale, more recent datasets than those that are publicly available online, the study focuses on the prediction of long-term renal illness with machine learning models [23]. The datasets were obtained from St. Paulo's Hospital in Ethiopia and comprise five classes: nocked, mild, moderate, severe, and ESRD; additionally, binary classes CKD were also applied. Since the majority of earlier studies concentrated on just two classes, it is challenging to propose a course of treatment because the stages determine the kind of care that should be administered [24].

4. Proposed Methodology

Machine learning is a branch of artificial intelligence that allows learners to process information without explicit programming. It centers on creating computer programmers with the ability to adapt to new information. It falls into one of two categories: Under supervision or not [25]. It all boils down to putting the right traits together to build frameworks that accomplish the right goals. Predictive clustering, parametric modeling, and multi-dimensional and multiclassification are a few examples of these activities [26]. The suggested methodology consists of three primary steps: data preparation, model training, and model selection as shown in Figure 1.

4.1 Dataset

Downloading a dataset from the Kaggle competition allows machine learning to be used to predict chronic kidney illness. The dataset included data from records belonging to 400 distinct patients. The list of 25 factors also includes the ages of the participants, bacterial species, albumin, which is the mineral potassium, creatinine in the serum, white blood cell count, and red blood cell count. In terms of categorization, appetite, packed cell volume, and blood glucose and urea levels, patients often display irregular and unpredictable patterns. According to research, the two main causes of chronic kidney disease (CKD) are diabetes and high blood pressure [27]. Anticipate elevated blood glucose levels as a normal byproduct of the harm that diabetes induces to our many.

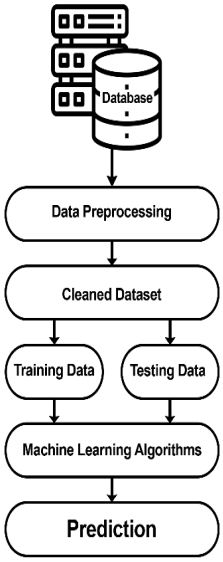

Figure 1. Proposed Model

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 400 entries, 0 to 399
Data columns (total 25 columns):

#	Column	Non-null Count	Datatype
0	age391	391 non-null	float64
1	blood_pressure	388 non-null	float64
2	specific_gravity	353 non-null	float64
3	albumin	354 non-null	float64
4	sugar	351 non-null	float64
5	red_blood_cells	248 non-null	object
6	pus_cell	335 non-null	object
7	pus_cell_clumps	396 non-null	object
8	bacteria	396 non-null	object
9	blood_glucose_random	356 non-null	float64
10	blood_urea	381 non-null	float64
11	serum_creatinine	383 non-null	float64
12	sodium	313 non-null	float64
13	potassium	312 non-null	float64
14	haemoglobin	348 non-null	float64
15	packed_cell_volume	330 non-null	object
16	white_blood_cell_count	295 non-null	object
17	red_blood_cell_count	270 non-null	object
18	hypertension	398 non-null	object
19	diabetes_mellitus	398 non-null	object
20	coronary_artery_disease	398 non-null	object
21	appetite	399 non-null	object
22	peda_edema	399 non-null	object
23	aanemia	399 non-null	object
24	class	400 non-null	object

Datatypes: float64(14), object(11)
Memory usage: 78.3+ KB

Figure 2. Features Listed in The Kidney Disease Dataset

4.2 Data Preprocessing

Two steps were taken in the data preparation process for this specific inquiry. We began by eliminating all the qualities (refer to Table I) that had data missing in excess of twenty percent. This fact directly leads to the exclusion of this specific collection of features from the investigation. We finished the duty of adding the principles that were missing from the remaining information obtained throughout the second step of the data preparation process. They must handle missing data by their distributions throughout the preparation stage [28]. They must handle missing data by their distributions throughout the preparation stage. To guarantee a respectable degree of accuracy, this is done. Little's MCAR test was used in this investigation to show that the missing values behaved erratically. This bias may be favorable

or negative, depending on what happened. A chi-square test of MCAR [29] is used to assess the methods of analysis that can be applied to fill in gaps in quantitative multivariate data. explores the hypothesis that there may be significant differences between the means of the different missing-value patterns. When data is transformed into a machine-readable format, the machine can interpret it more quickly and easily. A grouping of distinct data pieces is referred to as a "dataset" [30].

Fundamental qualities of data items can be identified and ensured more easily by using criteria like the volume or timing of an occurrence assured. The likelihood of missing values in the dataset is high; these can be computed or eliminated. When dealing with missing data, the values between the mode, median, and mean of the related variables can be utilized to fill out in the gaps [31]. The most popular approach to handling missing data is this one. Analysis cannot be done until numerical values with an object type are converted to float 64 bit values. The value that appears the most frequently in the attribute column is utilized in place of the null value when working with categorical attributes that have null values. Label encoding is a useful technique for converting category data into numerical attributes [32]. Giving each attribute value its integer value is required for this. This will directly result in the instant generation of an int data type. Each column's mean values are calculated beforehand, and any gaps in the corresponding attribute column are filled in with those values. Using the classifier function, the mean value for each column may be determined. The data must go through the procedures of testing, verification, and training after it has been updated and encoded [33]. Our algorithms learn from the data that we feed them, which gives them the knowledge they need to build a model. We use the dataset's validation section to improve the model and assess the precision of the various model fits we have made [34].

Table 1. Data Processing

Columns of data	Not null
age391	391 notnull
bloodpressure	388 notnull
specificgravity	353 notnull
albumin	354 notnull
sugar	351 notnull
redbloodcells	248 notnull
puscell	335 notnull
puscellclumps	396 notnull
bacteria	356 notnull
bloodglucoserandom	381 notnull
bloodurea	383 notnull
serumcreatinine	313 notnull
sodium	403 notnull
potassium	312 notnull
haemoglobin	348 notnull
packedcellvolume	329 notnull
whitebloodcellcount	294 notnull
redbloodcellcount	269 notnull
Hypertension	398 notnull
diabetesmellitus	398 notnull
coronaryarterydisease	398 notnull
Appetite	399 notnull
pedaedema	399 notnull
aanemia	399 notnull
class	400 notnull
Datatypes: float(14), objects(11)	
Memory consume: 78.3+ KB	

4.3 Machine Learning Models

The purpose of the study was to apply machine learning. techniques to forecast chronic renal disease. In this work, three tree machine learning methods were used: Random Forest, Support Vector Machine, and Decision Tree. The selection of algorithms was based on their popularity in the prediction of chronic renal disease and their classification performance on earlier research projects [35].

4.3.1 Random Forest

This type of ensemble learning uses many decision tree sets. Regression and classification are the two uses for it. This model consists of many decision trees, each of which produces a class target based on the target output's highest voting results [36]. Random Forest blends, The tree and a grove of unrelated trees are created using bagging and random feature selection. The collective prediction is more accurate than any single tree's. Test examples are filtered down through each tree once the forest is constructed, and the trees then forecast their respective classes [37].

4.3.2 Support vector machine (SVM)

This well-known and practical supervised machine-learning technique can be applied to learning, prediction, and classification. A collection of hyperplanes is constructed to sort every input into a high-dimensional data category. A discrete hyperplane is built in the signifier space of the training data, and compounds are categorized by the hyperplane's side [38]. The decision boundaries that divide the data points are called hyperplanes. To ascertain the location and direction of the hyperplane, data points that are closer to it are called support vectors. In the modern world, a large amount of data must be classified into more than two classes., SVMs have mostly been presented to address binary classification. However, in recent times, several academics have attempted to apply SVMs to multiclass classification. The two most common methods for solving multiclass issues with SVM are one-again-the-rest and one-versus-one [39]. In this study, one-versus-rest has been utilized. We employed OVR in conjunction with the SVM algorithm for multiclassification. Using this kind of technique, every class in the dataset is isolated from the other classes. One versus rest is also a suitable strategy using Linear SVC, since this study uses Linear SVC. The SVM pseudocode is displayed in Fig. 3.

Input: α: = 0 or α: = partially trained SVM, X and y,

1. C: =Approximate upper bound constant value
2. Loop: $\forall$ {X_i, yi} and {X_i, y_i}
3. do
4. Optimize α_i and α_j
5. End loop
6. Until no changes in a (or some resource constraint encounters)

Output: Hold merely support vectors (SV) (α_i >0)

Figure. 3 SVM pseudocode

4.3.3 Decision Trees

One of the most often used supervised machine-learning methods for classification is the decision tree (DT). By using sorted feature values to turn the data into a tree representation, Decision Trees solve the machine learning challenge. In a decision tree, each node indicates a feature that needs to be classified for an instance, and the class label to which the samples belong is shown by each leaf node. Since this model is predictive and makes use of observations about an item, it divides the dataset based on conditions using a tree structure. to calculate the goal value of instances [40]. While Figure 4 shows the binary class of chronic renal illness decision-making process, Figure 5 also shows the Decision Tree pseudocode.

Input: Data

1. Set the value of k
2. Loop: 1 to N // To get predicted class
   2.1. Calculate the distance Di (Euclidian/Cosine/Chebyshev) between data instance in training data and test data.
3. Increasingly arrange the computed distances (Di)
4. Populate the upper k results from the arranged list
5. Pick up the most frequent class from the list

Output: resultant class

Figure 4. Decision Tree pseudocode

5. Result and Discussion

The first and most important step in creating a precise machine-learning model is performance evaluation. To make sure the prediction model does well on unseen data and matches the dataset, it must be evaluated. The performance evaluation's goal is to calculate a model's generalization accuracy using out-of-sample data. One performance evaluation technique for analyzing and contrasting models that divides data into partitions is called cross-validation, or CV. Nine of the k folds, or equal-size subsamples, from the original dataset were utilized to instruct the model, and one of was used to validate or test it. After repetitions of this process, the average performance will be calculated. In this investigation, tenfold cross-validation was employed. Various performance evaluation criteria, such as f1-score, recall, accuracy, and precision, Specificity and sensitivity Could have been calculated. When the expected value and the actual value are both positive, this is known as an actual positive (TP) situation. When a data point's actual value and its anticipated value are both negative, it is considered to be truly negative (TN). False positives (FP): These are instances in which a data point's projected value is positive but its actual value was negative. False negativity (FN) are instances in which a data point's projected value is negative but its actual value is positive.

5.1 Accuracy

The capacity of the algorithm for classification to accurately anticipate the classes in dataset is implied by accuracy. It is a measurement of the degree to which the real or theoretical value and the expected value coincide [41]. Accuracy is typically expressed as the ratio of accurate forecasts to the total number of occurrences. Equation 1 displays the accuracy equation.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

5.2 Precision

The values that are accurately predicted from every expected value in the actual class are used to measure precision. The precision of the classifiers is measured by their inability to classify a negative example as positive. Since the macro average assigns each class the same weight, it is utilized for multiclass classification. Equation 2 displays the macro average precision equation.

$$Precision = \frac{TP + FP}{TP} \tag{2}$$

5.3 Recall

The rate at which positive values are accurately classified is called recall. What percentage of real positives are accurately classified is answered by recall. Equation 6 displays the equation of recollection. Since the models' recall value is determined using the macro average, the macro average recall is computed using the formula shown in Equation 3.

$$Recall = \frac{TP + FN}{TP} \tag{3}$$

5.4 F-Measure

The F-measure, which is the harmonic mean of recall and precision, is often referred to as the F1-score. Equation 4 displays the F1-score equation.

$$F1\_Score = \frac{Precision + Recall}{2 \times Precision \times Recall} \tag{4}$$

The strategies that yielded maximum degree of accuracy found in Tables' datasets were chosen based on these findings. Classifiers come in a variety of forms, such as KNN, Random Forests, XgBoost, Cat Boost, Ada Boost Classifier, Extra Trees Classifier, Gradient Boosting Classifier, and Stochastic Gradient Boosting. Once The relevance of the chosen features for each sort of prediction has been analyzed and concluded, a decision must be taken [42]. For each method, the standard deviation of the feature significance is calculated. The preferences of the algorithm for each of the several attributes are shown in Figure 5. When contrasting the random forest classifier with the decision tree classifier, It is evident that the former has fewest feature biases. Certain traits, such anemia-a, hunger, and pedal edoema, are overrepresented in the overall data, despite the distribution covering the whole spectrum of renal illness.

It was difficult to attain full accuracy without substituting values from a collaborative imputer rather than a constant for the missing data, which were lost completely at random [43]. Certain characteristics have a lesser correlation with medicinal value than others, depending on the patient's developmental stage. The training process has a significant impact on the accuracy of the model. The data set's distribution can be used to assist the distinct class separation that each of the chosen qualities, with the exception of serum creatinine, possesses [44].

Lastly, as Fig. 5 illustrates, certain trained models prioritize particular features over others when choosing the algorithm. When your account for the reasons that altered their nominal values, you have a lot more possibilities than only renal illness to think about. It was chosen because, as a result, using an extra tree classifier enables decision-makers to take into account several criteria rather than just one [45]. Upon applying diverse machine learning algorithms to the dataset, the following accuracy results were obtained. At 99%, the Extra Trees Classifier achieves the greatest accuracy.

Table 2. Accuracy Table

	Model	Score
8	Extra Trees Classifier	0.991667
4	Gradient Boosting Classifier	0.983333
5	Stochastic Gradient Boosting	0.983333
6	XgBoost	0.983333
7	Cat Boost	0.983333
1	Decision Tree Classifier	0.975000
2	Random Forest Classifier	0.975000
3	Ada Boost Classifier	0.975000
0	KNN	0.716667

Table 3. Classification Report

Classifier	Support	F1 Score	Recall	Precision
Extra Tree	72%	0.99	1.00	0.99
Random Forest	72%	0.98	0.96	0.97
Gradient Boost	79%	0.86	0.82	0.83
Stochastic Gradient	72%	0.98	0.97	0.94
Catboost Classifier	90%	0.98	0.97	0.96
Decision Tree	85%	0.92	0.97	0.95
KNN	72%	0.70	0.71	0.70

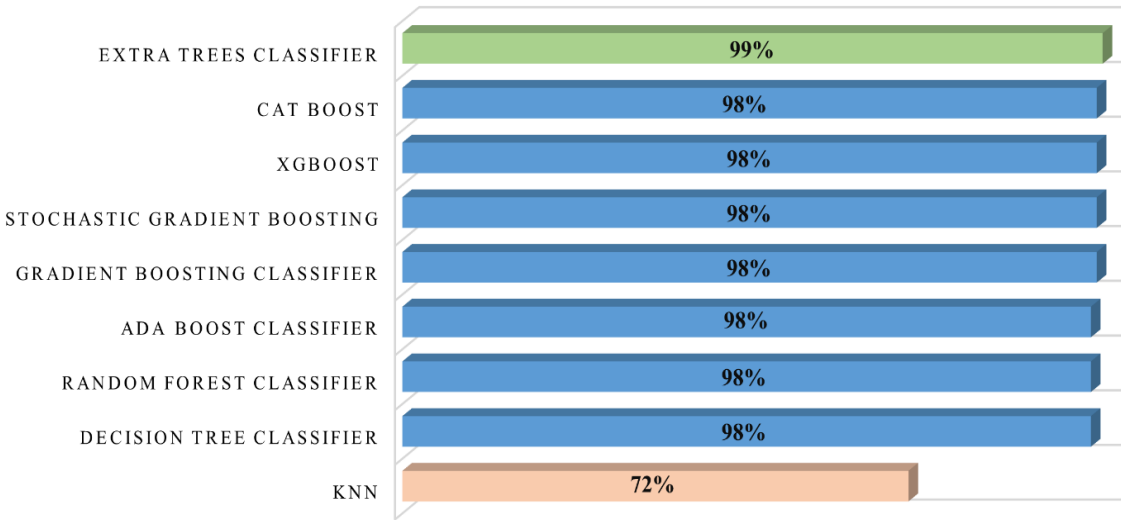

Figure 5. Accuracy comparison based on Chronic Renal Disease prediction

6. Conclusion and Future work

Various machine learning algorithms were applied to the chronic kidney disease dataset, and their performances were compared. The Extra Trees Classifier achieved the highest accuracy of 99% in predicting chronic kidney renal disease. The study concludes that machine learning approaches, particularly the Extra Trees Classifier, can be a useful tool for early diagnosis of chronic kidney disease, potentially leading to better patient outcomes through timely intervention. For future work, the study mentions that certain limitations need to be addressed, such as dealing with missing data more effectively using collaborative imputation rather than constant imputation. It also suggests incorporating additional relevant features and investigating more advanced machine learning techniques to further improve the prediction models. However, there is still scope for improvement by addressing data quality issues and exploring more sophisticated algorithms in future research.

References

1. Radhakrishnan J, Mohan S. KI Reports and World Kidney Day. Kidney Int Reports. 2017;2(2):125–6.
2. George C, Mogueo A, Okpechi I, Echoufo-Tcheugui JB, Kengne AP. Chronic kidney disease in low-income to middle-income countries: The case of increased screening. BMJ Glob Heal. 2017;2(2):1–10.
3. Ethiopia: kidney disease. https://www.worldlifeexpectancy.com/ethiopia-kidney-disease. Accessed 07 Feb 2020.
4. Stanifer JW, et al. The epidemiology of chronic kidney disease in sub-Saharan Africa: A systematic review and metaanalysis. Lancet Glob Heal. 2014;2(3): e174–81.
5. AbdElhafeez S, Bolignano D, D'Arrigo G, Dounousi E, Tripepi G, Zoccali C. Prevalence and burden of chronic kidney disease among the general population and high-risk groups in Africa: A systematic review. BMJ Open. 2018; 8:1.
6. Molla MD, et al. Assessment of serum electrolytes and kidney function test for screening of chronic kidney disease among Ethiopian Public Health Institute staff members, Addis Ababa, Ethiopia. BMC Nephrol. 2020;21(1):494.
7. Agrawal A, Agrawal H, Mittal S, Sharma M. Disease Prediction Using Machine Learning. SSRN Electron J. 2018; 5:6937–8.
8. Kumar V. Evaluation of computationally intelligent techniques for breast cancer diagnosis. Neural Comput Appl. 2021;33(8):3195–208.
9. J. C. Jakobsen, C. Gluud, J. Wetterslev, and P. Winkel, "When and how should multiple imputation be used for handling missing data in randomised clinical trials—A practical guide with flowcharts," *BMC Medical Research Methodology*, vol. 17, no. 1, p. 162, 2017.
10. S. D. Arasu and R. Thirumalaiselvi, "Review of chronic kidney disease based on data mining techniques," *International Journal of Applied Engineering Research*, vol. 12, no. 23, pp. 13498–13505, 2017.
11. Charleonnan A, Fufaung T, Niyomwong T, Chokchueypattanakit W, Suwannawach S, Ninchawee N. Predictive analytics for chronic kidney disease using machine learning techniques. Manag Innov Technol Int Conf MITiCON. 2016;80–83:2017.
12. Salekin A, Stankovic J. Detection of Chronic Kidney Disease and Selecting Important Predictive Attributes. In: Proc. - 2016 IEEE Int. Conf. Healthc. Informatics, ICHI 2016, pp. 262–270, 2016.
13. Tekale S, Shingavi P, Wandhekar S, Chatorikar A. Prediction of chronic kidney disease using machine learning algorithm. Disease. 2018;7(10):92–6.
14. Xiao J, et al. Comparison and development of machine learning tools in the prediction of chronic kidney disease progression. J Transl Med. 2019;17(1):1–13.
15. Almasoud M, Ward TE. Detection of chronic kidney disease using machine learning algorithms with least number of predictors. Int J Adv Computer. 2019;10(8):89–96.
16. Yashf SY. Risk Prediction Of Chronic Kidney Disease Using Machine Learning Algorithms. 2020.
17. S. Ramya and N. Radha, "Diagnosis of chronic kidney disease using machine learning algorithms," in *Proc. International Journal of Innovative Research in Computer and Communication Engineering*, 2016.
18. S. Vijayarani, S. Dhayanand, *et al.*, "Data mining classification algorithms for kidney disease prediction," *International Journal on Cybernetics & Informatics (IJCI)*, vol. 4, no. 4, pp. 13–25, 2015
19. Rady EA, Anwar AS. Informatics in Medicine Unlocked Prediction of kidney disease stages using data mining algorithms. Informatics Med. 2019;15(2018):100178.
20. Alsuhibany SA, et al. Ensemble of deep learning based clinical decision support system for chronic kidney disease diagnosis in medical internet of things environment. Comput Intell Neurosci. 2021;3:2021.
21. Poonia RC, et al. Intelligent Diagnostic Prediction and Classification Models for Detection of Kidney Disease. Healthcare. 2022;10:2.
22. L. Rubini. (2015). Early stage of chronic kidney disease UCI machine learning repository. [Online]. Available: http://archive.ics.uci.edu/ml/datasets/Chronic_Kidney_Disease
23. Estimated Glomerular Filtration Rate (eGFR). (Dec. 2015). [Online]. Available: https://www.kidney.org/atoz/content/gfr
24. Priyanka K, Science BC. Chronic kidney disease prediction based on naive Bayes technique. 2019. p. 1653–9.
25. Aqlan F, Markle R, Shamsan A. Data mining for chronic kidney disease prediction. 67th Annu Conf Expo Inst Ind Eng. 2017;2017:1789–94.
26. Subas A, Alickovic E, Kevric J. Diagnosis of chronic kidney disease by using random forest. IFMBEProc.2017;62(1):589–94.
27. Kapoor S, Verma R, Panda SN. Detecting kidney disease using Naïve bayes and decision tree in machine learning. Int J Innov Technol Explore Eng. 2019;9(1):498–501.
28. Vijayarani S, Dhayanand S. Data Mining Classification Algorithms for Kidney Disease Prediction. Int J Cybern Informatics. 2015;4(4):13–25.
29. Drall S, Drall GS, Singh S. Chronic kidney disease prediction using machine learning: a new approach bharat Bhushan Naib. Learn. 2014;8(278):278–87.

30. KadamVinay R, Soujanya KLS, Singh P. Disease prediction by using deep learning based on patient treatment history. Int J Recent Technol Eng. 2019;7(6):745–54.
31. Ramya S, Radha N. Diagnosis of Chronic Kidney Disease Using. pp. 812–820, 2016.
32. Osisanwo FY, Akinsola JET, Awodele O, Hinmikaiye JO, Olakanmi O, Akinjobi J. Supervised machine learning algorithms: classification and comparison. Int J Comput Trends Technol. 2017;48(3):128–38.
33. P. B. Jensen, L. J. Jensen, and S. Brunak, "Mining electronic health records: towards better research applications and clinical care," Nature Reviews Genetics, vol. 13, no. 6, pp. 395–405, 2012.
34. J. C. Ho, C. H. Lee, and J. Ghosh, "Septic shock prediction for patients with missing data," ACM Transactions on Management Information Systems (TMIS), vol. 5, no. 1, p. 1, 2014.
35. "Ictclas," http://ictclas.nlpir.org/
36. Y.-D. Zhang, X.-Q. Chen, T.-M. Zhan, Z.-Q. Jiao, Y. Sun, Z.-M. Chen, Y. Yao, L.-T. Fang, Y.-D. Lv, and S.-H. Wang, "Fractal dimension estimation for developing pathological brain detection system based on minkowski-bouligand method," IEEE Access, vol. 4, pp. 5937–5947, 2016.
37. S. Basu Roy, A. Teredesai, K. Zolfaghar, R. Liu, D. Hazel, S.Newman, and A. Marinez, "Dynamic hierarchical classification for patient risk-of readmission," in Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining. ACM, 2015, pp. 1691–1700.
38. Sak Haim, Andrew Senior, Franoise Beaufays, "Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition", 2014.D. W. Hosmer, S. Lemeshow, Applied Logistic Regression, Wiley Interscience, 2000.
39. C.J.C. Burges, "Simplified Support Vector Decision Rules", Proc. 13th Int'l Conf. Machine Learning, pp. 71-77, 1996.
40. Wang, Z., Chung, J. W., Jiang, X., Cui, Y., Wang, M., & Zheng, A. (2018). Machine learning-based prediction system for chronic kidney disease using associative classification technique. International Journal of Engineering & Technology, 7(4.36), 1161.
41. Sivasankar, E., Pradeep, R., & Sivanandham, S. (2019). Identification of important biomarkers for detection of chronic kidney disease using feature selection and classification algorithms. International Journal of Medical Engineering and Informatics, 11(4), 368–385. doi:10.1504/IJMEI.2019.104981
42. Rosso, R., Munaro, G., Salvetti, O., Colantonio, S., & Ciancitto, F. (2010, August). CHRONIOUS: an open, ubiquitous and adaptive chronic disease management platform for chronic obstructive pulmonary disease (COPD), chronic kidney disease (CKD) and renal insufficiency. In 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology (pp. 6850-6853). doi:10.1109/IEMBS.2010.5626451
43. Rustam, Z., Sudarsono, E., & Sarwinda, D. (2019, June). Random-forest (RF) and support vector machine (SVM) implementation for analysis of gene expression data in chronic kidney disease (CKD). IOP Conference Series. Materials Science and Engineering, 546(5), 052066. doi:10.1088/1757-899X/546/5/052066
44. E. Dritas and M. Trigka, "Machine learning techniques for chronic kidney risk prediction," Big Data and Cognitive Computing, pp. 1–15, 2022.
45. U. Ekanayake and D. Herath, "Chronic kidney disease prediction using machine learning methods," in Proc. 2020 Moratuwa Engineering Research Conference (MERCon), 2020, pp. 260–265.