

RESEARCH

Open Access



Chronic kidney disease prediction using machine learning techniques

Dibaba Adeba Debal^{1*} and Tilahun Melak Sitote²

*Correspondence:
dibabaadeba44@gmail.com

¹ Department of Information Science, College of Computing, Madda Walabu University, Robe, Ethiopia

² Department of Computer Science and Engineering, School of Electrical Engineering and Computing, Adama Science and Technology University, Adama, Ethiopia

Abstract

Goal three of the UN's Sustainable Development Goal is good health and well-being where it clearly emphasized that non-communicable diseases is emerging challenge. One of the objectives is to reduce premature mortality from non-communicable disease by third in 2030. Chronic kidney disease (CKD) is among the significant contributor to morbidity and mortality from non-communicable diseases that can affected 10–15% of the global population. Early and accurate detection of the stages of CKD is believed to be vital to minimize impacts of patient's health complications such as hypertension, anemia (low blood count), mineral bone disorder, poor nutritional health, acid base abnormalities, and neurological complications with timely intervention through appropriate medications. Various researches have been carried out using machine learning techniques on the detection of CKD at the premature stage. Their focus was not mainly on the specific stages prediction. In this study, both binary and multi classification for stage prediction have been carried out. The prediction models used include Random Forest (RF), Support Vector Machine (SVM) and Decision Tree (DT). Analysis of variance and recursive feature elimination using cross validation have been applied for feature selection. Evaluation of the models was done using tenfold cross-validation. The results from the experiments indicated that RF based on recursive feature elimination with cross validation has better performance than SVM and DT.

Keywords: Chronic Kidney Disease (CKD), Machine Learning, Random Forest (RF), Support Vector Machine (SVM)

Introduction

Chronic kidney disease (CKD) is non-communicable disease that has significantly contributed to morbidity, mortality and admission rate of patients worldwide [2]. It is quickly expanding and becoming one of the major causes of death all over the world. A report from 1990 to 2013 indicated that the global yearly life loss caused by CKD increased by 90% and it is the 13th leading cause of death in the world [1]. 850 million people throughout the world are likely to have kidney diseases from different factors [3]. According to the report of world kidney day of 2019, at least 2.4 million people die every year due to kidney related disease. Currently, it is the 6th fastest-growing cause of death worldwide. CKD is becoming a challenging public health problem with increasing prevalence worldwide. Its burden is even higher in low-income countries where detection,

prevention and treatment remain low [2]. Kidney disease is serious public health problem in Ethiopia affecting hundreds of thousands of people irrespective of age, sex [4]. The lack of safe water, appropriate diet, and physical activities is believed have contributed. Additionally, communities living in rural area have limited knowledge about the CKD. According to WHO report of 2017, the number of deaths in Ethiopia due to kidney disease was 4,875. It is 0.77% of total deaths that has ranked the country 138th in the world. The age-adjusted death rate is 8.46 per 100,000 of the population and the death rate increased to 12.70 per 100,000 that has ranked the country 109 in 2018 [3].

National kidney foundation classifies stages of CKD into five based on the abnormal kidney function and reduced Glomerular Filtration Rate (GFR), which measures a level of kidney function,. The mildest stage (stage 1 and stage 2) is known with only a few symptoms and stage 5 is considered as end-stage or kidney failure. The Renal Replacement Therapy (RRT) cost for total kidney failure is very expensive. The treatment is not also available in most developing countries like Ethiopia. As a result, the management of kidney failure and its complications is very difficult in developing countries due to shortage of facilities, physicians, and the high cost to get the treatment [4, 5]. Hence, early detection of CKD is very essential to minimize the economic burden and maximize the effectiveness of treatments [6]. Predictive analysis using machine learning techniques can be helpful through an early detection of CKD for efficient and timely interventions [7]. In this study, Random Forest (RF), Support Vector Machine (SVM) and Decision Tree (DT) have been used to detect CKD. Most of previous researches focused on two classes, which make treatment recommendations difficult because the type of treatment to be given is based on the severity of CKD.

Related works

Different machine-learning techniques have been used for effective classification of chronic kidney disease from patients' data.

Charleonnann et al. [8] did comparison of the predictive models such as K-nearest neighbors (KNN), support vector machine (SVM), logistic regression (LR), and decision tree (DT) on Indians Chronic Kidney Disease (CKD) dataset in order to select best classifier for predicting chronic kidney disease. They have identified that SVM has the highest classification accuracy of 98.3% and highest sensitivity of 0.99.

Salekin and Stankovic [9] did evaluation of classifiers such as K-NN, RF and ANN on a dataset of 400. Wrapper feature selection were implemented and five features were selected for model construction in the study. The highest classification accuracy is 98% by RF and a RMSE of 0.11. S. Tekale et al. [10] worked on "Prediction of Chronic Kidney Disease Using Machine Learning Algorithm" with a dataset consists of 400 instances and 14 features. They have used decision tree and support vector machine. The dataset has been preprocessed and the number of features has been reduced from 25 to 14. SVM is stated as a better model with an accuracy of 96.75%.

Xiao et al. [11] proposed prediction of chronic kidney disease progression using logistic regression, Elastic Net, lasso regression, ridge regression, support vector machine, random forest, XGBoost, neural network and k-nearest neighbor and compared the models based on their performance. They have used 551 patients' history data with proteinuria with 18 features and classified the outcome as mild, moderate,

severe. They have concluded that Logistic regression performed better with AUC of 0.873, sensitivity and specificity of 0.83 and 0.82, respectively.

Mohammed and Beshah [13] conducted their research on developing a self-learning knowledge-based system for diagnosis and treatment of the first three stages of chronic kidney have been conducted using machine learning. A small number of data have been used in this research and they have developed prototype which enables the patient to query KBS to see the delivery of advice. They used decision tree in order to generate the rules. The overall performance of the prototype has been stated as 91% accurate.

Priyanka et al. [12] carried out chronic kidney disease prediction through naive bayes. They have tested using other algorithms such as KNN (K-Nearest Neighbor Algorithm), SVM (Support Vector Machines), Decision tree, and ANN (Artificial Neural Network) and they have got Naïve Bayes with better accuracy of 94.6% when compared to other algorithms.

Almasoud and Ward [13] aimed in their work to test the ability of machine learning algorithms for the prediction of chronic kidney disease using subset of features. They used Pearson correlation, ANOVA, and Cramer's V test to select predictive features. They have done modeling using LR, SVM, RF, and GB machine learning algorithms. Finally, they concluded that Gradient Boosting has the highest accuracy with an F-measure of 99.1.

Yashfi [14] proposed to predict the risk of CKD using machine learning algorithms by analyzing the data of CKD patients. Random Forest and Artificial Neural Network have been used. They have extracted 20 out of 25 features and applied RF and ANN. RF has been identified with the highest accuracy of 97.12%.

Rady and Anwar [15] carried out the comparison of Probabilistic Neural Networks (PNN), Multilayer Perceptron (MLP), Support Vector Machine (SVM), and Radial Basis Function (RBF) algorithms to predict kidney disease stages. The researchers conducted their research on a small size dataset and few numbers of features. The result of this paper shows that the Probabilistic Neural Networks algorithm gives the highest overall classification accuracy percentage of 96.7%.

Alsuhibany et al. [16] presented ensemble of deep learning based clinical decision support systems (EDL-CDSS) for CKD diagnosis in the IoT environment. The presented technique involves Adaptive Synthetic (ADASYN) technique for outlier detection process and employed ensemble of three models, namely, deep belief network (DBN), kernel extreme learning machine (KELM), and convolutional neural network with gated recurrent unit (CNN-GRU).

Quasi-oppositional butterfly optimization algorithm (QOBOA) technique is also employed in the study for hyperparameter tuning of DBN and CNN-GRU. The researchers have concluded that EDL-CDSS method has the capability of proficiently detecting the presence of CKD in the IoT environment.

Poonia et al. [17] employed Various machine learning algorithms, including k-nearest neighbors algorithm (KNN), artificial neural networks (ANN), support vector machines (SVM), naive bayes (NB), and Logistic Regression as well as Re-cursive Feature Elimination (RFE) and Chi-Square test feature-selection techniques. Publicly available dataset of healthy and kidney disease patients were used to build and analyze prediction models.

The study found that a logistic regression-based prediction model with optimal features chosen using the Chi-Square technique had the highest accuracy of 98.75%.

Vinod [18] carried out the assessment of seven supervised machine learning algorithms namely K-Nearest Neighbor, Decision Tree, Support vector Machine, Random Forest, Neural Network, Naïve Bayes and Logistic Regression to find the most suitable model for BCD prediction based on different performance evaluation. Finally, the result showed that k-NN is the best performer on the BCD dataset with 97% accuracy.

The above reviews indicates that several studies have been conducted on chronic kidney disease prediction using machine-learning techniques. There are various parameters which play important role in improving model performance like dataset size, quality of dataset and the time dataset collected. This study focuses on chronic kidney disease prediction using machine learning models based on the dataset with big size and recent than online available dataset collected from St. Paulo's Hospital in Ethiopia with five classes: notckd, mild, moderate, severe, and ESRD and binary classes: ckd and notckd by applying machine-learning models. Most previously conducted researches focused on two classes, which make treatment recommendations difficult because the type of treatment to be given is based on the stages. Table 1 below shows the summary of some related works.

Table 1 Summary of some related works

No.	Author	Technique applied	Claimed outcome	Draw back
1	Salekin and Stankovic [9]	K-NN, RF, and NN, Wrapper approach and Embedded approach	Detection F1-score of RF 99.8	Small dataset size with missing values was used; Severity level prediction was not included
2	Tekale et al. [10]	DT and SVM	DT and SVM with an accuracy of 91.75 and 96.75 respectively	Dataset size need to increased, Severity level prediction was not included. Only to classifiers result compared
3	Priyanka et al. [12]	NB, KNN, SVM, DT, and ANN. NB	NB, KNN, SVM, DT, and ANN. NB accuracy is 94.6%	Small size dataset. No stages prediction. Feature extraction was not carried out and classification accuracy needs improvement
4	Yashfi [14]	RF and ANN	RF and ANN with an accuracy of 97.12% and 94.5%	Small size dataset and no stages prediction
5	Rady and Anwar [15]	PNN, MLP, SVM, RBF	PNN, MLP, SVM, RBF accuracy of 96.7, 60.7, 87, and 51.5 respectively	Small data size. the algorithms, which is not appropriate for small dataset size were used
6	Alsuhibany et al. [16]	EDL-CDSS technique, ADASYN based outlier detection and QOBOA based hyperparameter tuning	The proposed EDL-CDSS method has depicted the other approaches with the superior sensy and specy of 0.9680 and 0.9702 compared to ACO, FNC, KELM, CNN-GRU, DBN, DT, MLP and D-ACO	The size of benchmark dataset is small. IoT and benchmark data integration is not clearly explained. The study focus only on binary classification (ckd or not-ckd)
7	Poonia et al. [17]	KNN, ANN, SVM, NB and LR with Chi-Square and RFE	KNN, ANN, SVM, NB, LR accuracy of 66.25%,65%,97.5%,95%,97.5% respectively	Small size dataset and no stages prediction. Accuracy needs improvement

Materials and method

Data source and description

The data source for this study is St. Paulo’s Hospital. It is the second-largest public hospital in Ethiopia which admits large number of patients with chronic diseases. There are dialysis treatment and kidney transplant center in the hospital. As it has been shown in Table 2, the dataset for this study is patients’ records of chronic kidney disease from patients admitted to the renal ward during 2018 to 2019. Some of them were obtained from the same patient history data at different times of different stages. To prepare the dataset and understand features, interviews of domain experts have been conducted. The dataset contains 1718 instances with 19 features where 12 are numeric and 7 are nominal. As the detail have been shown in Table 3, the features in the dataset include Age, Gender, blood pressure, specific gravity, chloride, sodium, potassium, blood urine nitrogen, serum creatinine, hemoglobin, red blood cell count, white blood cell count, mean cell volume, platelet count, hypertension, diabetic mel-litus, anemia and heart disease. When we see multi class distribution, 441 (25.67%) instances are end-stage renal disease stage or stage five, 399 (23.22%) are at a severe stage or stage four, 354 (20.61%) are at a moderate stage or stage three, 248 (14.44%) are at a mild stage or stage two, and 276 (16.07%) have no chronic kidney disease or normal. The class distribution for binary class is 1442 (83.93%) ckd (stage 1 to 5) and 276 (16.07%) notckd. The binary-class distribution is imbalanced. Oversampling data resampling technique have been used to balance the value of the minority class with the value of the majority class. After employing the resampling technique, the total size of binary class dataset, become 2888.

Preprocessing

Real world data is often inconsistent which can affect the performances of models. Preprocessing the data before it is fed into classifiers is vital part of developing

Table 2 Dataset description

Description	Values
Data source	St. Paul’s Hospital, Addis Ababa, Ethiopia
Period	2018 to 2019
No. of instances	1718
No. of features	19
Class	Both binary and multi class (5 stages)
End-stage renal disease stage (stage five)	25.67%
Severe stage (stage four)	23.22%
Moderate stage (stage three)	20.61%
Mild stage (stage two)	14.44
Percentage of negative samples	16.07%

Table 3 Features description

Symbols	Features full name	Type	Class	Missing values in %
Age	Age	Numeric	Predictor	0
Gender	Gender	Nominal	Predictor	0
Bp	Blood pressure	Numeric	Predictor	0.058207
Sg	Specific gravity	Nominal	Predictor	0.058207
Chl	Chloride	Numeric	Predictor	0.116414
Sod	Sodium	Numeric	Predictor	0.232829
Pot	Potassium	Numeric	Predictor	0.116414
Bun	Blood Urea Nitrogen	Numeric	Predictor	0.232829
Scr	Serum Creatinine	Numeric	Predictor	0.058207
Hgb	Hemoglobin	Numeric	Predictor	0.232829
Rbcc	Red blood cell count	Numeric	Predictor	0.232829
Wbcc	White blood cell count	Numeric	Predictor	0.232829
Mcv	Mean cell volume	Numeric	Predictor	6.111758
Pltc	Platelet count	Numeric	Predictor	7.275902
Htn	Hypertension	Nominal	Predictor	0
Dm	Diabetes Mellitus	Nominal	Predictor	0
Ane	Anemia	Nominal	Predictor	0
Hd	Heart disease	Nominal	Predictor	0
ckd_status	ckd_status	Nominal	Target	0

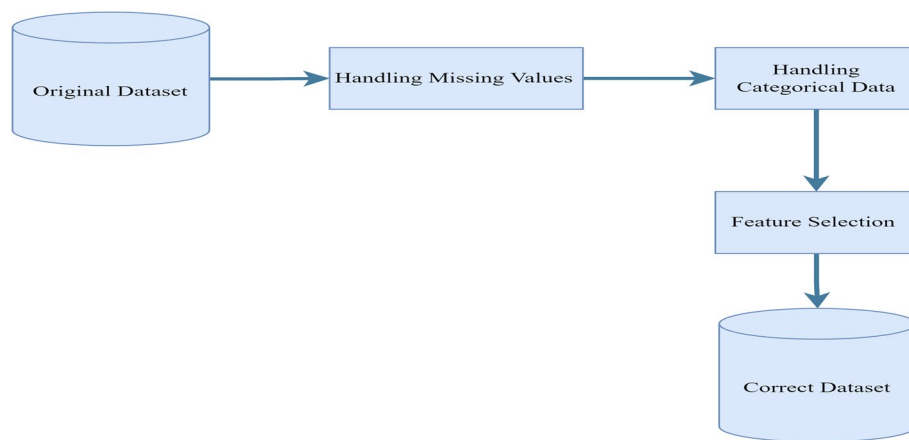


Fig. 1 Chronic kidney disease dataset preprocessing steps

machine-learning model. Similarly, the dataset for this study contains missing values that needs to be handled appropriately. It has to also be in a suitable format for modeling. Hence, pre-processing has been conducted as it has been shown in Fig. 1.

Cleaning Noisy Data: removing outlier and smoothening noisy data is an important part of preprocessing. Outliers are values that lies away from the range of the rest of the values. In clinical data, outliers may arise from the natural variance of data. The potential outliers are the data points that fall above $Q3 + 1.5(IQR)$ and below $Q1 - 1.5(IQR)$, where $Q1$ is the first quartile, $Q3$ is the third quartile, and $IQR = Q3 - Q1$ [19].

Handling Missing Values: data is not always available (or missed) due to equipment malfunction, inconsistent with other recorded data and thus deleted, not entered into

the database due to misunderstanding, some data may not be considered important at the time of entry.

Patient data often has missing diagnostic test results that would help to predict the likelihood of diagnoses or predict treatment's effectiveness [20]. The missing values have an impact on the performance of the prediction model. There are several ways of handling missing values including dropping missing values and filling missing values. Sometimes missing values are ignored if they are small percentage i.e., if missing data under 10%. But it is not considered healthy for the model because the missing value can be an important feature contributing to the model development. Sometimes the missing values can be also replaced by zero, which will bring no change to the model. To handle these missing values the mean, an average of the observed features used in this study, because the missing features are numeric and mean imputation is better for numerical missing values.

Handling Categorical Data: In this step, data has been transformed into the required format. The nominal data converted into numerical data of the form 0 and 1. For instance, 'Gender' has the nominal value that can be labeled as 0-for female and 1-for male. After preprocessing the data then the resultant CSV file comprises all the integer and float values for different CKD related features.

Normalization: It is important to scale numerical features before fitting to any models, as scaling is mandatory for some techniques such as nearest neighbors, SVMs, and deep learning [21]. There are different techniques of scaling and in this study Z-score normalization (or zero-mean normalization) have been used. The values for a feature are normalized based on the mean and standard deviation. It is as follows:

$$z = \frac{x - \mu}{\sigma} \quad (1)$$

where z is Z-score, x is feature value, μ is mean value and σ is standard deviation.

Feature selection

Identify subset of relevant predictive features is important for quality result [22]. Feature selection is the process of selecting most important predictive features to use them as input for models. It is important preprocessing step to deal with the problem of high dimensionality. Hence, the main aim of feature selection is to select the subset of features that are relevant and independent of each other for training the model [23]. Similarly, feature selection is crucial to develop chronic kidney disease predictive model. This reduces the dimensionality and complexity of the data and makes the model be faster, more effective and accurate. Hence, feature selection algorithm have been used to select relevant features after the construction of the dataset.

Filter, wrappers, or embedded techniques are widely used for feature selection in different clinical datasets including chronic kidney disease. A filter method is a technique that is independent of a classification algorithm and uses general characteristics of data for evaluating and selecting relevant features. The filter method works independently of the learning algorithm to remove irrelevant features and analyses properties of a dataset to chooses relevant features [24]. It is widely used approach due to its less complex in nature. With wrapper method, relevant features are selected using the classification

algorithm. It is better than filter feature selection technique in terms of accuracy. However, it requires higher processing time. In this study, univariate feature selection method from filter methods have been selected because it is fast, efficient and scalable. Recursive feature elimination with cross-validation (RFECV) has been used from wrapper feature selection method.

Univariate Feature Selection (UFS): This method is popular, simplistic and fastest feature selection method used in healthcare dataset. It considers each feature separately to determine the strength of the relationship of the feature with the dependent variable. It is fast, scalable and independent of classifier. Different options are there for univariate algorithms such as Pearson correlation, information gain, chi-square, ANOVA (Analysis of Variance). In this study, feature selection was done using the ANOVA as shown in Eq. 2.

$$F = \frac{MST}{MSE} \quad (2)$$

where F is ANOVA coefficient, MST is mean sum of squares of treatment and MSE is mean sum of squares error.

Recursive Feature Elimination with Cross-Validation (RFECV): An optimization algorithm to develop a trained machine-learning model with relevant and selected features by repeatedly eliminating irrelevant features. It repetitively creates the model, keeps aside the worst performing feature at each iteration, and builds the next model with the remaining features until the features are completed to select best subset of features [25]. It eliminates the redundant and weak feature whose deletion least affects the training and keeps the independent and strong feature to improve the generalization performance of the model [26]. This method uses the iterative procedure for feature ranking and to find out the features that have been evaluated as most important. Because this technique work interacting with a machine learning model, it first builds the model on the entire set of features and ranked the feature according to its importance.

Machine learning models

The aim of the study was to predict chronic kidney disease using machine-learning techniques. Three machine learning algorithms; Random Forest, Support Vector Machine and Decision Tree have been used in this study. The algorithms were selected based on their popularity in chronic kidney disease prediction and their performance of classification on previous research works [12, 27–33].

Random Forest: Random Forest is ensemble learning that consists of several collections of decision trees. It is used for both classification and regression. This model comprises of a number of decision trees and outputs the class target that is the highest voting results of the target output by each tree [28]. Random Forest uses both bagging and random feature selection to build the tree and creates an uncorrelated forest of trees. Its prediction by the group is more accurate than that of any individual tree. After it builds the forest, test instances are permeated down through each tree and trees make their respective prediction of class [33]. The Random Forest pseudocode is shown in Fig. 2.

Support vector machine (SVM): Support Vector Machine is one of the prominent and convenient supervised machine-learning algorithm that can be used for

Input: Test data

1. Predict and store the outcome of each randomly created decision trees (D Tree's) on given test data
2. Compute the total votes for individual class
3. Declare majority class as the final outcome class

Output: Final predicted class

Fig. 2 Random forest pseudocode [18]

Input: $\alpha = 0$ or $\alpha =$ partially trained SVM, X and y,

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

Output: Hold merely support vectors (SV) ($\alpha_i > 0$)

Fig. 3 SVM pseudocode [18]

classification, learning and prediction. A set of hyperplanes are built to classify all input in high dimensional data. A discrete hyperplane is created in the signifier space of the training data and compounds are classified based on the side of the hyperplane [30]. Hyperplanes are decision boundaries that separate the data points. Support vectors are data points that are closer to the hyperplane and determine the position and orientation of the hyperplane. SVMs have been mainly proposed to deal with binary classification, but nowadays many researchers have tried to apply it to multiclass classification because there are a huge amount of data to be classified into more than two classes in today's world. SVM solves multiclass problems through the two most popular approaches; one-versus-rest and one-vs-one. In this study, one-versus-rest has been used. For multi classification, we used OVR with the SVM algorithm. This method separates each class from the rest of the classes in the dataset. Besides, because Linear SVC is used in this study, one vs rest is an appropriate method with Linear SVC. The pseudocode of SVM have been shown in Fig. 3.

Decision Tree (DT): It is one of the most popular supervised machine-learning algorithms that can be used for classification. Decision Tree solves the problem of machine learning by transforming the data into a tree representation through sorted feature values. Each node in a decision tree denotes features in an instance to be classified, and each leaf node represents a class label the instances belong to. This model uses a tree structure to split the dataset based on the condition as a predictive model that maps observations about an item to make a decision on the target value of instances [34]. Decision Tree pseudocode is shown in Fig. 4 and decision making in binary class of chronic kidney disease have been shown in Fig. 5.

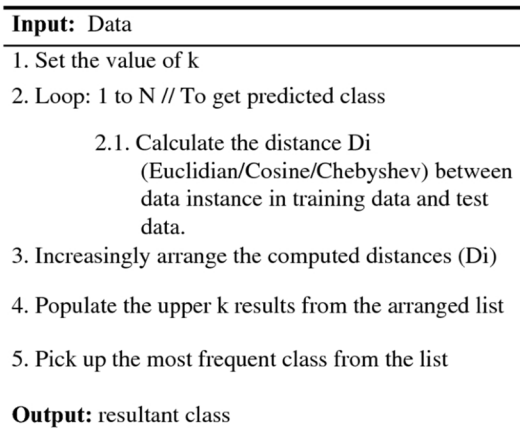


Fig. 4 Decision tree pseudocode [18]

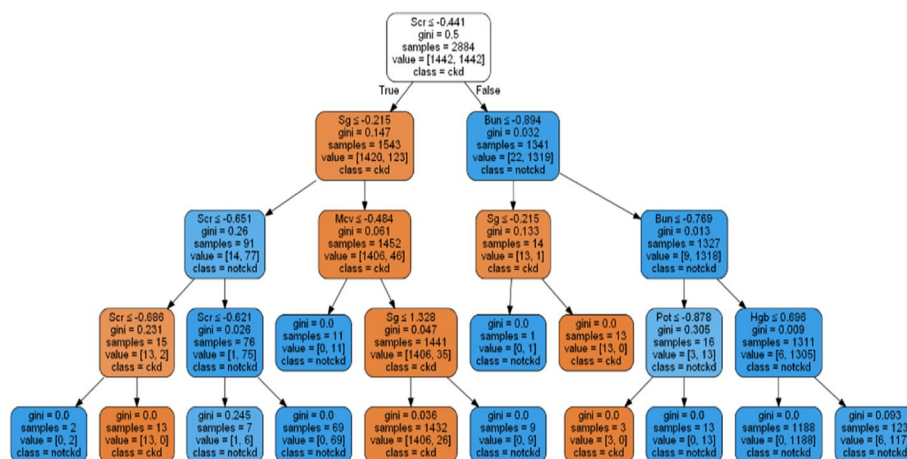


Fig. 5 Decision making in binary class of chronic kidney disease

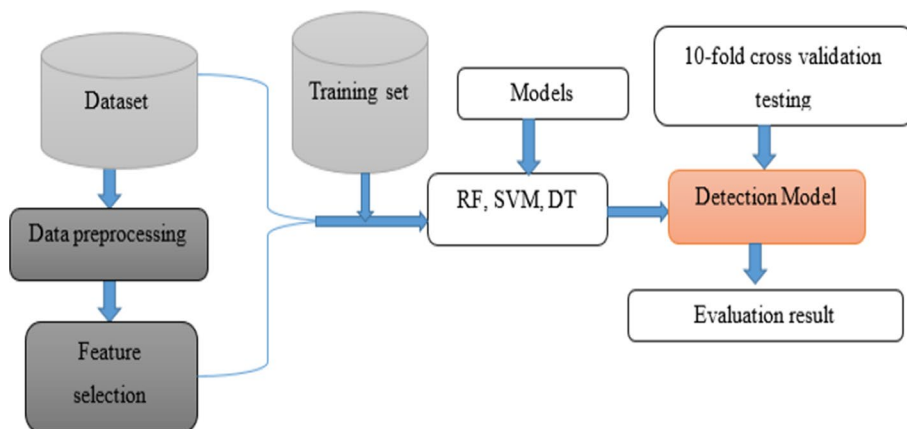


Fig. 6 Model building flow diagram

Figure 6 shows the flow of model building using three machine-learning algorithms with tenfold cross validation. The machine learning models were developed for both multiclass and binary classification. The best performing model was selected as best machine learning model from the three algorithms for each classification.

Prediction model evaluation

Performance evaluation is the critical step of developing an accurate machine-learning model. Prediction model shall to be evaluated to ensure that the model fits the dataset and work well on unseen data. The aim of the performance evaluation is to estimate the generalization accuracy of a model on unseen/out-of-sample data. Cross-Validation (CV) is one of the performance evaluation methods for evaluating and comparing models by dividing data into partitions. The original dataset was partitioned into k equal size subsamples called folds: nine used to train a model and one used to test or validate the model. This process repeated k times and the average performance will be taken. Tenfold cross-validation have been used in this study. Different performance evaluation metrics including accuracy, precision, recall, f1-score, sensitivity, specificity have been computed.

- True positive (TP): are the condition when both actual value and predicted value are positive.
- True negative (TN): are the condition when both the actual value of the data point and the predicted are negative.
- False positive (FP): These are the cases when the actual value of the data point was negative and the predicted is positive.
- False negative (FN): are the cases when the actual value of the data point is positive and the predicted is negative.

Accuracy

Accuracy implies the ability of the classification algorithm to predict the classes of the dataset correctly. It is the measure of how close or near the predicted value is to the actual or theoretical value [35]. Generally, accuracy is the measure of the ratio of correct predictions over the total number of instances. The equation of accuracy is shown in Eq. 3.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (3)$$

Precision

Precision measure the true values correctly predicted from the total predicted values in the actual class. Precision quantifies the ability of the classifiers to not label a negative example as positive. The equation of precision is shown in Eq. 4.

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

Macro average is used for multiclass classification because it gives equal weight for each class. The equation of macro average precision is shown in Eq. 5.

$$Precision_macro = \frac{1}{|C|} \sum_{i=1}^{|C|} \frac{TP_i}{TP_i + FP_i} \quad (5)$$

Recall

Recall measure the rate of positive values that are correctly classified. Recall answers the question of what proportion of actual positives are correctly classified. The equation of recall is shown in Eq. 6.

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

Since the macro average is used in order to compute the recall value of the models, macro average recall is calculated as follows (Eq. 7).

$$Recall_macro = \frac{1}{|C|} \sum_{i=1}^{|C|} \frac{TP_i}{TP_i + FN_i} \quad (7)$$

F-measure

F-measure is also called F1-score is the harmonic mean between recall and precision. The equation of F1-score is shown in Eq. 8.

$$F1_score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (8)$$

The macro average of F1_score is calculated as follows (Eq. 9).

$$F1_score_macro = 2 * \frac{Precision_macro * Recall_macro}{Precision_macro + Recall_macro} \quad (9)$$

Sensitivity

Sensitivity is also called True Positive Rate. Sensitivity is the mean proportion of actual true positives that are correctly identified [36]. The equation of sensitivity is shown in Eq. 10.

$$Sensitivity = \frac{TP}{TP + FN} \quad (10)$$

Specificity

Specificity is also called True Negative Rate. It is used to measure the fraction of negative values that are correctly classified. The equation of sensitivity is shown in Eq. 11.

$$Specificity = \frac{TN}{TN + FP} \quad (11)$$

Results and discussions

Feature selection

The features selection process using the two methods; UFS and RFECV resulted two different sets of features. The resulted subset of features have been used in the training of RF, SVM, and DT. The selected features for both five-class and binary-class are different because recursive feature elimination with cross-validation can automatically eliminate less predictive features iteratively depending on the model. Table 4 shows the number of selected features of binary-class and five-class that set the size of the dataset respectively.

Evaluation results

The Experiment has been carried out based on two feature selection methods and three classifiers for both binary and five-class classification for 18 models. Training and testing on these models have been executed using tenfold cross-validation. In tenfold cross-validation, the dataset is partitioned randomly into ten equal size sets. Training the models was then done by using 10–1 folds and test using one remaining fold. The process is iterative for each fold. The obtained results are presented in binary and five-class classification models. Modeling was first carried out for both binary and five class classification using preprocessed dataset without applying feature selection methods. Then, modeling has been experimented by applying the two feature selection methods as it has been discussed in the following sections.

Binary classification models evaluation results

These classification models were built using the two-class dataset that was converted from the five-class dataset. The target classes are notckd and ckd. The models are trained and tested using tenfold CV and other performance evaluation metrics of CV. As it has been discussed previously, modeling was first conducted on the preprocessed dataset without applying feature selection methods. Then, feature selection techniques have been implemented to select the most predictive features. Feature selection was implemented using UFS and RFECV. The performance measures of each test for RF, SVM, and DT models before feature selection and after feature selection method are presented in Table 5.

An accuracy of 99.8% resulted in the RF with RFECV model with selected 8 features is the highest. The result of models before applying feature selection is graphically shown in Fig. 7.

Table 4 Feature selection results

Feature selection methods	Machine learning algorithm	Selected features (Binary)	Selected features (five-Class)
RFECV	RF	8	9
	SVM	9	10
	DT	16	7
UFS (ANOVA)	RF	14	14
	SVM	14	14
	DT	14	14

Table 5 Binary classification without and with feature selection

Models	No of features	Acc	Prec	Rec	F1	Sen	Spec
RF	18	99.7	99.9	99.5	99.7	99.5	99.9
RF with RFECV	8	99.8	99.9	99.7	99.8	99.7	99.9
RF with UFS	14	99.7	99.9	99.5	99.7	99.5	99.9
SVM	18	96.9	99.1	94.7	96.8	94.7	99.1
SVM with RFECV	9	95.5	98.7	92.2	95.3	92.2	98.8
SVM with UFS	14	96.7	99.4	97.7	98.6	97.7	99.4
DT	18	98.5	99.4	97.6	98.5	97.6	99.4
DT with RFECV	16	98.6	99.4	97.7	98.6	97.7	99.4
DT with UFS	14	98.5	99.4	97.6	98.5	97.6	99.4

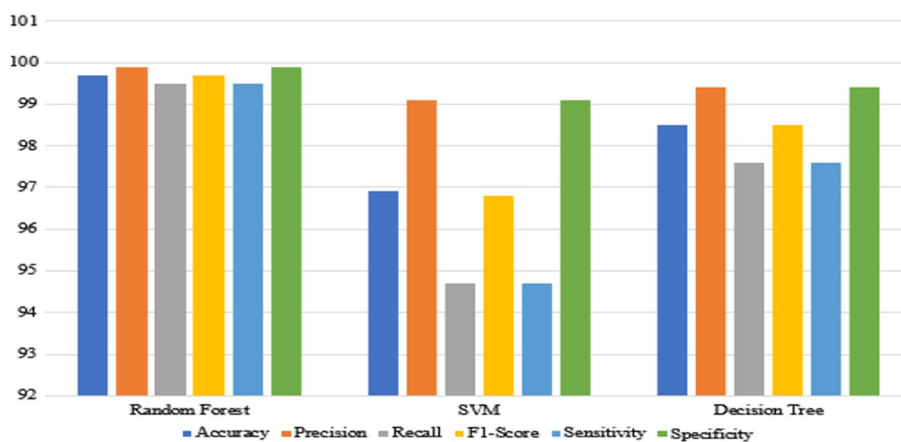


Fig. 7 Binary class classification without feature selection

Furtherly, we have carried out hyperparameter optimization using grid search using cross validation on SVM for binary dataset. This has been done without feature selection. The performance has been significantly improved to 99.83% which is almost the same with the highest RF with RFECV. Another experiment on the binary classification is with Extreme Gradient Boosting (XGBoost). It is a powerful machine learning algorithm that can be used to solve classification and regression problems. The performance is 98.96% which is not better than RF with RFECV’s performance.

Multiclass classification models evaluation results

The multiclass models were similarly built using the preprocessed five-class dataset. The models are trained and tested using tenfold CV and evaluated with other performance evaluation metrics. The performance metrics result of each trained model; RF, SVM, and DT have been presented for without and with feature selection method as shown in Table 6. Models were first trained and tested with all features without applying feature selection methods and then we apply the feature selection methods.

Table 6 shows the CV performance metrics results for three classifiers of multiclass dataset before and after applying feature selection. The accuracy is 79% from RF with RFECV with selected 9 features. The result of models after applying feature selection

Table 6 Multiclass classification results without and with feature selection

Models	No. of features	Acc	Prec	Rec	F1	Sen	Spec
RF	18	78.3	79.5	77.4	77.3	77.4	94.5
RF with RFECV	9	79.0	79.8	78.1	77.9	78.1	94.7
RF with UFS	14	77.6	78.9	76.6	76.7	76.7	94.3
SVM	18	63.0	60.9	61.5	59.9	61.6	90.6
SVM with RFECV	10	62.2	58.7	60.7	56.5	60.8	90.5
SVM with UFS	14	61.5	59.2	60.5	58.4	60.5	90.2
DT	18	77.5	79.6	76.4	76.2	76.5	94.3
DT with RFECV	7	78.0	80	76.9	76.6	76.9	94.4
DT with UFS	14	77.9	79.8	76.7	76.4	76.7	94.4

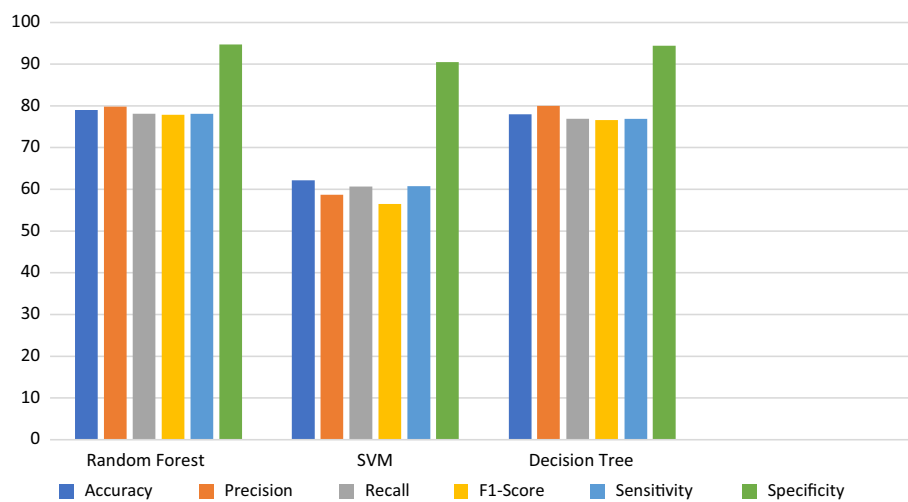


Fig. 8 Multiclass classification results after applying RFECV

is graphically shown in Fig. 8. Similarly, we have carried out hyperparameter optimization using grid search using cross validation on SVM for multiclass dataset. This has been done without feature selection. The performance has been significantly improved to 78.78% which is not better than the highest performing model RF with RFECV. Another experiment on the multiclass classification is with Extreme Gradient Boosting (XGBoost). It is a powerful machine learning algorithm that can be used to solve classification and regression problems. The performance is 82.56% which is better than RF with RFECV’s performance.

Discussions

Chronic kidney disease is a global health threat and becoming silent killer in Ethiopia [6]. Many people die or suffer severely by the disease mainly due to lack of awareness about the disease and inability to detect early. Thus, early prediction of chronic kidney disease is believed to be helpful to slow the progress of the disease. Machine learning plays a vital role in early disease identification and prediction. It supports the decision of medical experts by enabling them to diagnose the disease fast and accurately.

In this study, chronic kidney disease prediction has been carried out using machine learning techniques. The datasets consists of 19 features with numerical and nominal values along with the class to which each instance belongs. The dataset had missing values and these missing values were handled at the preprocessing step. After preprocessing, we have prepared two datasets for two tasks; binary classification and multiclass classification that has five-classes. The dataset comprises 1718 instances. The binary class dataset was created by converting the five-class dataset to two class labels; notckd and ckd. All classes except notckd converted to ckd in the binary class dataset. After preparing the dataset to binary classification, it has been observed that the class labels were imbalanced and we applied the data resampling technique to balance the dataset.

Three machine-learning models and two feature selection methods were used. The model evaluation was done using tenfold cross-validation and other performance evaluation metrics such as precision, recall, F1-score, sensitivity, and specificity. The confusion matrix was used to show the correctly and incorrectly classified classes to evaluate the performance of the model. Initially, we applied machine learning classifiers without feature selection for both binary class and five classes. The machine learning models used in this study are Random Forest, Support Vector Machine, and Decision tree. Then, feature selection techniques were applied along with the models in order to select predictive. RFECV and UFS were the two feature selection methods applied to select the relevant features. UFS is the filtering method, it works independently of the machine learning model and RFECV is the wrapper method that works depending on the machine-learning algorithm. RFECV is also automatic feature selection method that can select features without specifying the number of features to be selected. The number of selected features is different from model to model as it has been shown in the results. There were features that has been frequently selected in each feature sets, which indicates that they are the most predictive features and have strong predictive relation to the class. The most frequently selected features using both feature selection methods in all models are serum creatinine (Scr), blood urine Nitrogen (Bun), Hemoglobin (Hgb), and Specific Gravity (Sg). Pltc, Rbcc, Wbcc, Mcv, Dm, and Htn are the next most frequently selected features.

Several related studies have been conducted to predict chronic kidney disease using machine learning and other techniques globally. However, there were very limited works in the context of Ethiopia. The focuses of global studies were mainly on binary classification. There were very limited attempt on multiclass classification. Additionally, the size of the datasets used were small which makes them susceptible for overfitting and difficult for comparison.

Salekin and Stankovic [9] proposed the method of detecting chronic kidney disease using K-NN, RF and NN, analysed the characteristics of 24 features, and sorted their predictability. They have used the dataset of 400 instances comprising 250 ckd and 150 notckd. They have used five features for final model construction and they have evaluated the performance of their model using tenfold cross-validation. Compared to the proposed solution in this study, this work used small size dataset and only focus on binary classification (ckd and notckd). Even though dataset used in the previous work and dataset used in proposed study are not the same, the proposed model in this study has higher performance compared to the previous work.

In the study conducted by Almasoud and Ward [13] Logistic regression, support vector machines, random forest, and gradient boosting algorithms and feature selection methods such as the ANOVA test, the Pearson's correlation, and the Cramer's V test were implemented to detect chronic kidney disease. A dataset with 400 instances has been used. Compared to the proposed model performance of binary class classification, the performance of the model in this work is inferior. The kidney disease stages prediction model was developed by Rady and Anwar using PNN, MLP, SVM and RBF [15]. They evaluated and compared the performance of the models with accuracy. They reported accuracy accuracy result were 96.7% for PNN, 87% for RBF, 60.7% for SVM and 51.5% for MLP. They used dataset of 361 instances with 25 features including the target class. The total data used in this work has been the data of patients with ckd and then they calculated the eGFR to identify stages of the disease. PNN with an accuracy of 96.7% and RBF with an accuracy of 87% developed by Rady and Anwar [15] shows better performance than the model with best performance in this study which is RF based on RFECV with accuracy of 79% even though models in [15] and two models build in this study are different. However, the SVM model built in this study perform better than the SVM model built by Rady and Anwar [15]

In this study, proposed models were evaluated using tenfold cross-validation and other performance evaluation metrics such as precision, recall, *f1_scor*, sensitivity, and specificity. Four machine-learning models; RF, SVM, DT and Extreme Gradient Boosting (XGBoost) have been implemented before applying feature selection. SVM resulted the highest accuracy of 99.8% for binary. Its accuracy is 78.78% for five classes. RF resulted accuracy of 99.7% for binary class and accuracy of 78.3% for five-class. XGBoost resulted accuracy of 98.96% for binary class. The accuracy of XGBoost is 82.56% for five-class which is the highest. DT resulted accuracy of 98.5% for binary class and accuracy of 77.5% for five-class. Then, the two feature selection methods were applied to the two datasets. Both SVM and RF with RFECV resulted the highest accuracy for the binary class. XGBoost has the highest accuracy of 82.56% for five-class. The result is promising and we believe that it can be deployed to support medical experts to identify the disease fast and accurate. Thus, SVM, RF with RFECV and XGBoost is recommended in our study based on its accuracy, *f1score*, and other performance evaluation for binary class and five-class classifications.

Conclusion

Early prediction is very crucial for both experts and patients to prevent and slow down the progress of chronic kidney disease to kidney failure. In this study three machine-learning models RF, SV, DT, and two feature selection methods RFECV and UFS were used to build proposed models. The evaluation of models were done using tenfold cross-validation. First, the four machine learning algorithms were applied to original datasets with all 19 features. Applying the models on the original dataset, we have got the highest accuracy with RF, SVM, and XGBoost. The accuracy was 99.8% for the binary class and 82.56% for five-class. DT produced lowest performance compared to RF. RF also produced the highest *f1_score* values. SVM and RF with RFECV produced the highest accuracy of 99.8%for binary class. XGBoost has 82.56% accuracy for five-class datasets which is the highest. Hencewe believe that multi classification work was very important

to know the stages of the disease and suggest needed treatments for the patients in order to save their lives.

Future works

This study used a supervised machine-learning algorithm, feature selection methods to select the best subset features to develop the models. It is better to see the difference in performance results using unsupervised or deep learning algorithms models. The proposed model supports the experts to give the fast decision, it is better to make it a mobile-based system that enables the experts to follow the status of the patients and help the patients to use the system to know their status.

Acknowledgements

Not applicable.

Author contributions

The development of the basic research questions, identifying the problems and selecting appropriate machine learning algorithms, data collection, data analysis, interpretation, and critical review of the paper have been done by DA. The edition of the overall progress of the work was supported and guided by TM. Both authors read and approved the final manuscript.

Funding

Not applicable.

Availability of data and materials

The data to conduct this research was collected from the patient history data attends their treatment or died during a period 2018 to 2019 from St. Paulo's Hospital. There are no restrictions on the availability of data and the authors are willing to provide the code as well.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 31 January 2022 Accepted: 24 October 2022

Published online: 20 November 2022

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, Echouffo-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 meta-analysis. *Lancet Glob Heal*. 2014;2(3):e174–81.
5. AbdElhafeez S, Bolognani 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. Charleonnann 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.
9. 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.
10. Tekale S, Shingavi P, Wandhekar S, Chatorikar A. Prediction of chronic kidney disease using machine learning algorithm. *Disease*. 2018;7(10):92–6.
11. 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.

12. Priyanka K, Science BC. Chronic kidney disease prediction based on naive Bayes technique. 2019. p. 1653–9.
13. 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.
14. Yashfi SY. Risk Prediction Of Chronic Kidney Disease Using Machine Learning Algorithms. 2020.
15. Rady EA, Anwar AS. Informatics in Medicine Unlocked Prediction of kidney disease stages using data mining algorithms. *Informatics Med*. 2019;15(2018):100178.
16. Alshuibany 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.
17. Poonia RC, et al. Intelligent Diagnostic Prediction and Classification Models for Detection of Kidney Disease. *Healthcare*. 2022;10:2.
18. Kumar V. Evaluation of computationally intelligent techniques for breast cancer diagnosis. *Neural Comput Appl*. 2021;33(8):3195–208.
19. Jasim A, Kaky M. Intelligent systems approach for classification and management of by. 2017.
20. Saar-tsechansky M, Provost F. Handling Missing Values when Applying Classification Models. vol. 1, 2007.
21. Data Preparation for Statistical Modeling and Machine Learning. <https://www.feature-ranking.com/tutorials/machine-learning-tutorials/data-preparation-for-machine-learning/>. Accessed 12 Oct 2020.
22. Oliver T. Machine Learning For Absolute Beginners. 2017.
23. ZarPhyu T, Oo NN. Performance comparison of feature selection methods. *MATEC Web Conf*. 2016;42:2–5.
24. Koshy S. Feature selection for improving multi-label classification using MEKA. *Res J*. 2017;12(24):14774–82.
25. Vidhya A. Introduction to Feature Selection methods with an example (or how to select the right variables?). <https://www.analyticsvidhya.com/blog/2016/12/introduction-to-feature-selection-methods-with-an-example-or-how-to-select-the-right-variables/>. Accessed 24 Mar 2020.
26. Misra P, Yadav AS. Improving the classification accuracy using recursive feature elimination with cross-validation. *Int J Emerg Technol*. 2020;11(3):659–65.
27. Aqlan F, Markle R, Shamsan A. Data mining for chronic kidney disease prediction. *67th Annu Conf Expo Inst Ind Eng*. 2017;2017:1789–94.
28. Subas A, Alickovic E, Kevric J. Diagnosis of chronic kidney disease by using random forest. *IFMBE Proc*. 2017;62(1):589–94.
29. Kapoor S, Verma R, Panda SN. Detecting kidney disease using Naïve bayes and decision tree in machine learning. *Int J Innov Technol Explor Eng*. 2019;9(1):498–501.
30. Vijayarani S, Dhayanand S. Data Mining Classification Algorithms for Kidney Disease Prediction. *Int J Cybern Informatics*. 2015;4(4):13–25.
31. 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.
32. 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.
33. Ramya S, Radha N. Diagnosis of Chronic Kidney Disease Using. pp. 812–820, 2016.
34. 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.
35. Acharya A. Comparative Study of Machine Learning Algorithms for Heart Disease Prediction 2017.
36. Amirgaliyev Y. Analysis of chronic kidney disease dataset by applying machine learning methods. In: 2018 IEEE 12th International Conference Application Information Communication Technology, pp. 1–4, 2010.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
