Applied Medical Informatics
Original Research Vol. 45, No. 2/2023, pp: 58-69

Accuracy Prediction of Classification and Forecast using WEKA Tool by Example:Chronic Kidney Disease

Gayathri SAKHAMURI and Krupanidhi SREERAMA*

Department of Biotechnology, Vignan's Foundation for Science, Technology and Research, Vadlamudi 522213 A. P. India
Emails: gayathrisakhamuri2277@gmail.com; *sknbirac@gmail.com
* Author to whom correspondence should be addressed; Mobile Phone: 9440984668

Received: May 17, 2023/Accepted: June 25, 2023 / Published online: July 1, 2023

Abstract
Introduction: Renal failure due to kidney disease can be avoided with early diagnosis. Disease markers able to anticipate renal failure at an asymptomatic stage and thus, the onset of chronic kidney disease in a human subject can be predicted using, for example, data mining techniques. The present study focuses on building a decision tree and predicting the accuracy of machine learning classifiers to forecast kidney disease using the CKD dataset. *Methods*: The dataset in the current study includes information from 400 samples (instances) and 25 attributes retrieved from the freely available UCI machine learning repository. The accuracy of prediction of classifiers was conducted with the WEKA software tool using 14 algorithms. The performance evaluation of the models was done with accuracy, precision, recall and F-measure. *Results*: The lowest performance was given by Stacking and Vote classifiers. Moderate performance evaluation was observed for Logistic, Naïve Bayes, Random Tree, and Voted Perceptron. The best performance was observed for Random Forest, Multilayer Perceptron, Logit Boost, J48, Decision Table, Bagging, PART, and SMO. The following two were found to be statistically significant: Random Forest and Multilayer Perceptron . *Conclusion*: The decision tree could successively depict the contribution of serum creatine, pedal edema, diabetes, hemoglobin, and specific gravity of blood in tracing the prevalence of CKD in a prospective patient.

Keywords: Chronic Kidney Disease; Risk factors; Forecast; WEKA tool; Decision Tree

Introduction

Among the organ systems prevailing in the human body, the following are involved in vital physiological functions: neuro-muscular, cardio-vascular, gastrointestinal and renal systems. The above-listed organ systems are continuously functioning throughout our life to maintain physiological homeostasis. The renal system plays an important role in keeping the blood osmolarity, ion concentration, pH, fluid volume, blood pressure, release of hormones, and ultimately excretion of waste and toxins. The crux of the renal system is the kidney. Hence, the healthy state of the kidney is warranted. Due to age-related issues, life style, drinking water quality, food habits, and late-onset of disorders such as diabetes, hypertension, chronic inflammation, and frequent microbial infections render illness to kidneys which can be managed to prolong the life of affected subjects with prior knowledge on the causative features of chronic kidney diseases (CKD). The estimated glomerular filtration rate (eGFR) of 90-60 ml/min/1.73 m^2 is found to be the normal output of a healthy kidney whose values reduce either due to the rise in the levels of creatinine in the blood or proteinuria or co-morbidities [1].

Most people with kidney diseases living in rural areas and overcrowded homes have poor access to nephrological care in India[2]. In Andhra Pradesh (India), the Uddanam region,located in north-central districts comprising more than 100 villages with a population of ~150,000, had a high incidence of kidney diseases. People in this region suffer from chronic kidney disease unknown etiology (CKDu), with an incidence from 40% to 60% [3]. Upon screening more than 100,000 people in Uddanam area, 13% of the population in the Srikakulam district had a serum creatinine level of 1.2 mg/dL or higher [1]. Further, it was shown that men are more affected by CKD than women. In the adjoining state located at north of Andhra Pradesh, namely Odisha (India), significant CKDu prevalence rates were also reported [4].

Chronic kidney disease is a severe public concern due to persistent renal function impairments. A global prevalence of 13.4% and an annual mortality rate of 1.2 million due to CKD are reported [5]. Furthermore, 40% of CKD patients in the Marathwada region belonging to Maharashtra (India) had almost no recognized etiology. In Canacona, located in the western part of India in South Goa (India), it was reported a high frequency of CKDu [6].

One must rely on community-based data to assess the prevalence of CKD at an early stage because early CKD is often asymptomatic. According to the techniques employed and the sporadic endemic populations examined, the prevalence of CKD ranged from 4.5% to 17.5%, as indicated by Evans and Taal [7]. The estimated prevalence of CKD was found to be increased in the USA from 1988-1994 to 1999-2004, the same was noticed based on the patient data relating to albuminuria and glomerular filtration rate (GFR), classified under five stages and also attributed to the increased prevalence of diabetes and hypertension [8].

Health professionals sometimes rely on machine learning (ML) technology for prognosis, prophylaxis, and preventive measures. The dependence on trained health workers in the community is progressively increasing as the knowledge pervading among the prospective patients on ML classifier algorithm is scarce [9].

Machine learning tools have many public health applications. Osman and Sabit [10] used the Chi-squared automatic interaction detection algorithm on vaccination. The data on preterm bradycardia of 30 infants compiled at the Royal Hobart Hospital, Australia involved with 3591 h of ECG, was evaluated using ANNs (Artificial neural networks) and tested on new infants, which gave a mean value of 0.63 AUC (Area under the ROC Curve, where ROC is receiver operating characteristic) to predict bradycardia [11].The various ML (Machine learning)-based tools used in Machine learning Kidney Disease Diagnosis (MLKDD) were summarized and categorized by Qezelbash-Chamak et al. [12].

Machine learning tools have important application areas for health intervention programs [13] as data analysis and pattern recognition are components of ML [14]. Data mining in CKD has a major impact on unfolding hidden information from the extensive patients' medical treatment dataset that hospitals acquire to decipher the symptomatic data, facilitating an accurate therapeutic approach at an early stage. Thus, data mining is an essential pre-requisite for discovering the hidden and unnoticed knowledge in a medical dataset. One can forecast, categorize, classify, and cluster data using data mining techniques. These processes would analyze how the chosen algorithm will handle a training set that contains a number of attributes (disease markers) and outcomes. Nevertheless, ML techniques are ideally adopted across the globe for the recurrence prediction of the disease in the population and the accuracy of diagnosis prediction [6, 8, 15-19]. As the computerized hospital dataset grows exponentially, more and more ML algorithms are integrated into healthcare applications [20]. The ML capabilities in healthcare applications particularly relating to the cardiovascular system are classified into five classes: (i) digital imaging, (ii) electrocardiography, (iii) in-hospital monitoring, (iv) mobile and wearable technology and (v) precision medicine. It has been inferred that integration of such algorithms in daily practice could annually save up to US$600 per individual in the USA [21]. Park et al.[22] adopted the SHapley additive explanation method to determine the prediction of clinical features relating to ten diseases (namely malaria, hepatitis, liver cirrhosis, acute pancreatitis, acute myocardial infarction, unstable angina, acute pyelonephritis, end-stage renal disease, infectious colitis, and pulmonary tuberculosis) and proposed predictive relationship models with the clinical features and the laboratory tests. In the healthcare sector and the industrial and small-medium enterprises, the ML applications are widely being authenticated and

reaffirmed through a structural model comprising supply chain reengineering capabilities and also supply chain agility using partial least square-based equations [23]. Thus, ML classifiers are the state-of-the-art tools having societal, industrial, and healthcare research applications.

Socio-economic awareness of kidney diseases and related risk factors, disease markers, physiological symptoms, frequency of occurrence etc., are yet to be made familiar to the local community and community health workers. The Primary Health Centres, District level hospitals and corporate hospitals in India are treating the end-stage renal disease (ESRD) with renal replacement therapy (RRT), the strategies having significant family and national burdens. Known risk factors for CKD include age, sex, race, ethnicity, family history, drug usage, smoking, and low socioe-economic level. Synchronous CKD diseases, include diabetes and hypertension that are either conventionally or in-conventionally linked to CKD [15]. The awareness of the risk factors, disease markers and lifestyle habits are essential in averting the onset of kidney diseases. To create such an awareness, the machine learning tools, datasets on the disease and accuracy prediction analysis will provide the information to the community and village-level health workers far earlier than the onset of end-stage renal disease. Considering all the above-listed facts, in the present study, the prediction accuracy of the different chosen classifiers on CKD dataset was undertaken.

Material and Method

We used a dataset on CKD that is publically available at UCI Machine Learning Repository (https://archive.ics.uci.edu/ml/datasets/Chronic_Kidney_Disease) [24]. Four hundred persons in the dataset were classified with- namely CKD and without CKD- notCKD. Twenty-five disease-related data was collected for each patient (Table1).

Table 1. Attributes in kidney disease dataset as given in UCI machine learning repository.

Symbol	Description
age	Age
bp	Blood Pressure
sg	Specific gravity
al	Albumin
su	Sugar
rbc	Red Blood Cells
pc	Pus Cells
pcc	Pus Cell Clumps
ba	Bacteria
bgr	Blood Glucose Random
bu	Blood Urea
sc	Serum Creatinine

Symbol	Description
Sod	Sodium
Pot	Potassium
Hemo	Hemoglobin
Pcv	Packed Cell Volume
Wc	White Blood Cell Count
Rc	Red Blood Cell Count
Htn	Hypertension
Dm	Diabetes Mellitus
Cad	Coronary Artery Disease
Appet	Appetite
Pe	PedalEdema
Ane	Anemia
Class	Class

The "Explorer" environment of WEKA software was used as it contains data mining algorithms [25].

The CKD dataset used in the present study was unbalanced and pre-processed in WEKA software tool by choosing the icons in the series: Filters-Unsupervised-Attribute-Normalize. The used methods workflow is shown in Figure 1.

The dataset is saved in ARFF (Attribute related file format) on the desktop and uploaded in the WEKA tool. We used the normalized option under pre-process tab to ensure and standardize the format in the dataset. In the next step, we accessed the classifier folders: Bayes, Functions, Meta, Rules and Trees (Figure 2).

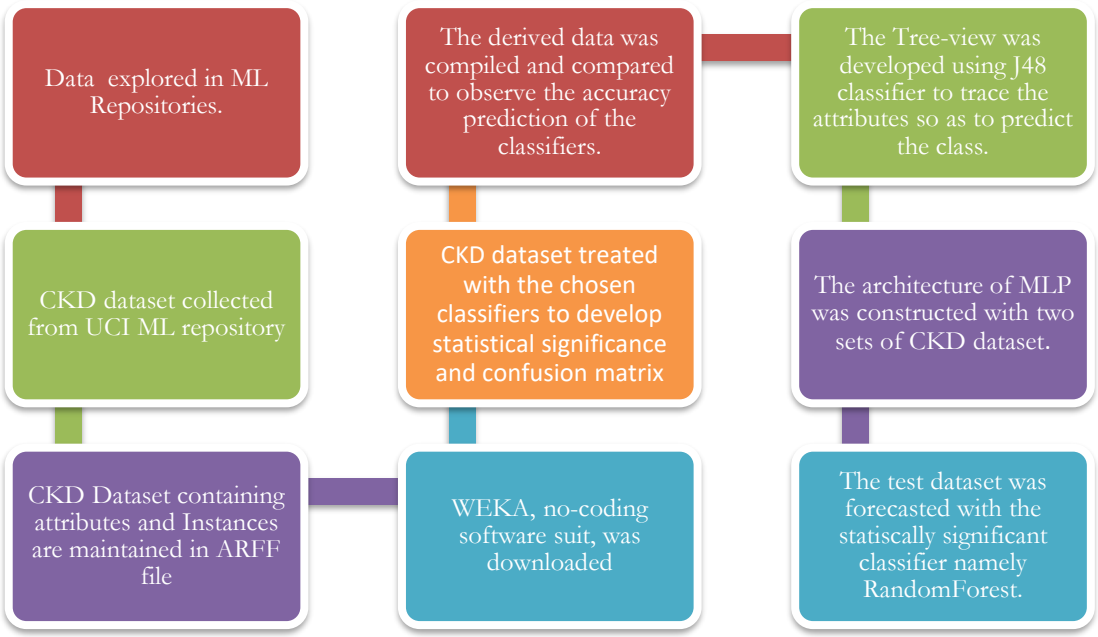

Figure 1. Flowchart representing the sequence of work done in the present investigation

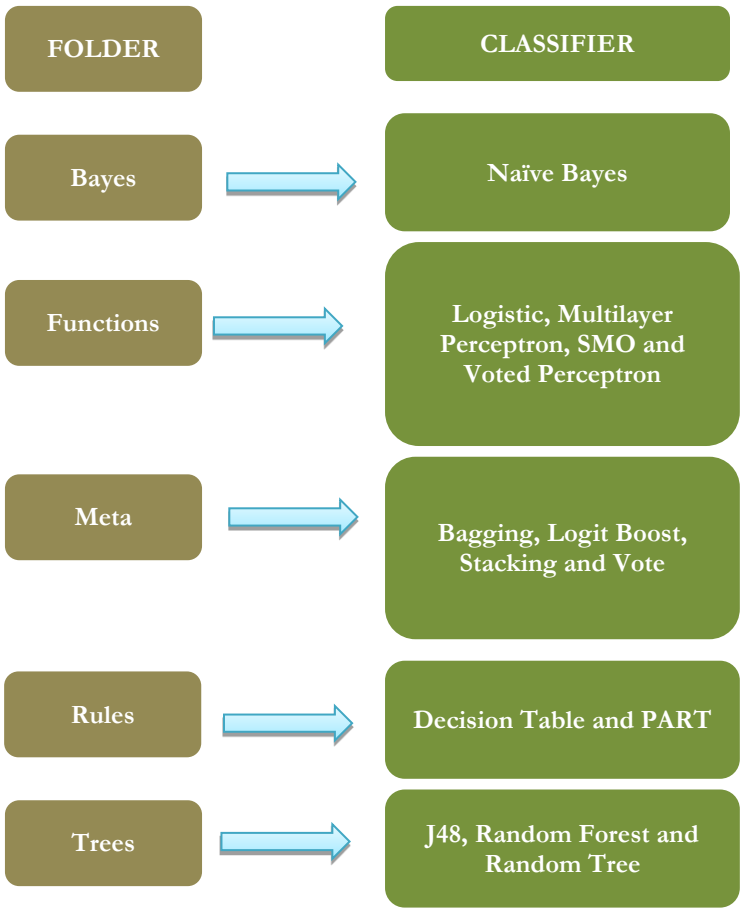

Figure 2. The classifiers in each of the foders chosen from WEKA tool

We used five classifier folders in our study (Figure 2). Each one is included with various algorithmic models. They are unique in testing, validating and developing models using the dataset They range from statistical-based learning models to neural network models to multiple training subsets to association rules to build correlation models and ultimately to decision trees with nodes representing a test.

The tree view was developed using J48 for the present data on CKD. The forecast using edited test dataset was performed with the classifier with the highest accuracy.

The metrics used to characterize the algorithms are presented in Table 2.

Table 2. Metrics for performance evaluation of the classifiers

Metric	Formula
Precision	[TP/(TP+FP)]× 100
Recall	[TP/(TP+FN)]× 100
Accuracy	[(TP+TN)/(TP+TN+FP +FN)]× 100
F-Measure	2 × (precision×recall)/(precision + recall)

Results

The accuracy prediction and statistical parameters obtained while using the classifiers were all tabulated in Table 3. Random Forest classifier was shown with a high kappa statistics value of 1.000 and also with the specificity of 100 (Table 4).

Table 3. The derived statistical values upon submission of the Kidney Disease dataset from UCI to the chosen classifiers in WEKA tool

Algorithm	Time (s)	Acc(%)	InAcc (%)	Kappa stat	MAE	RMSE	RAE(%)	RRSE(%)
NaïveBayes	0.01	96	4	0.9164	0.0429	0.1798	9.1489	37.1374
Logistic	0.09	95.75	4.25	0.9102	0.044	0.2062	9.3896	42.5854
MultilayerPerceptron	1.7	99.75	0.25	0.9947	0.0085	0.0622	1.8073	12.8559
SMO	0.01	97.75	2.25	0.9526	0.0225	0.15	4.7982	30.9838
Voted Perceptron	0.03	62.52	37.5	0	0.375	0.6124	79.9705	126.491
Bagging	0.08	98.75	1.25	0.9734	0.0545	0.1204	11.6266	24.8732
LogitBoost	0.07	99.75	0.25	0.9947	0.0199	0.0641	5.0858	17.1324
Stacking	0.01	62.5	37.5	0	0.4689	0.4841	100	100
Vote	0	62.5	37.5	0	0.4689	0.4841	100	100
DecisionTable	0.13	99	1	0.9786	0.1815	0.2507	38.7083	51.7879
PART	0.03	98.5	1.5	0.9680	0.0293	0.1014	6.2454	20.9492
J48	0.04	99	1	0.9786	0.0225	0.0807	4.243	13.2433
Random Forest	0.08	100	0	1.0000	0.0414	0.0844	8.8189	17.4439
RandomTree	0	95.5	4.5	0.905	0.045	0.1677	9.6037	34.6333

Time (s)= time taken to build the model (sec); Acc = Correct classified instances; InAcc (%) = In-correctly classified instances; Kappa stat = kappa statistics; MAE = Mean absolute error; RMSE = Root mean squared error; RAE = Relative absolute error (%); RRSE = Root relative squared error (%)

Random Forest stood out in classifying the instances (patient samples) correctly with 100% score, whereas, Multilayer Perceptron, Logit Boost, J48 and Decision Table classified instances correctly in the range of 99% to 99.75% (Table 3). True positive rate values indicating the correctly classified instances as correct instances were obtained for the classifiers in the order given herewith: Random Forest>Multilayer Perceptron>Logit Boost>Decision Table > J48 > Bagging > PART > SMO> Logistic >Random Tree> Naïve Bayes.

Table 4 presents the specificity of the classifiers considering the data given in the confusion matrix Table 5.

Table 4. The specificity of the classifiers used for accuracy prediction on CKD dataset was calculated considering the data given in the confusion matrix Table 5

Classifiers	Specificity = TN/(TN+FP) (%)
Naïve Bayes	90.36
Logistic	94.23
Multilayer Perceptron	99.33
SMO	95.54
Voted Perceptron	94.30
Bagging	97.38
Logit Boost	98.03
Stacking	0
Vote	0
Decision Table	100
PART	98
J48	99.32
Random Forest	100
Random Tree	96.10

Table 5. Confusion matrix for the correctly and in-correctly classified instances of CKD dataset using classifiers in WEKA software tool

Classifiers		True values	
		(CKD) Positive	(NotCKD) Negative
NaïveBayes	**CKD**	234	16
	notCKD	0	150
Logistic	**CKD**	241	9
	notCKD	3	147
Multilayer Perceptron	**CKD**	249	1
	notCKD	0	150
SMO	**CKD**	243	7
	notCKD	0	150
Voted Perceptron	**CKD**	241	9
	notCKD	1	149
Bagging	**CKD**	246	4
	notCKD	1	149
Logit Boost	**CKD**	247	3
	notCKD	0	150
Stacking	**CKD**	250	0
	notCKD	150	0
VOTE	**CKD**	250	0
	notCKD	150	0
Decision Table	**CKD**	250	0
	notCKD	5	145
PART	**CKD**	247	3
	notCKD	3	147
J48	**CKD**	249	1
	notCKD	3	147
Random Forest	**CKD**	250	0
	notCKD	0	150
Random Tree	**CDK**	244	6
	notCKD	2	148

The tree view was developed using J48 classifier to represent the possible presence of disease markers which ultimately diagnose the suspected human subject as CKD or notCKD (Figure 3). The tree is comprised of root node, branches, internal nodes, and leaf node. The root node generated in the tree view in J48 classifier with 'sc' (serum creatinine). The root node gave two branches, of which one branch led to the internal node 'pe' (pedal edema) (sc<=1.2) and the other (sc>1.2) ended with leaf node depicting the condition CKD. The internal node 'pe' again yielded two branches, the branch with pedal edema terminated with CKD and the branch with no pedal edema led to the internal node. Internal node with 'dm' (diabetes mellitus) contained two more branches namely 'yes' that lead to CKD and branch 'no' 'dm' led to another internal node hemo (haemoglobin). In continuation, the node 'hemo' was divided into two branches where the branch with hemo<=12.9 led to the leaf node indicating CKD and the other with hemo>12.9 was directed to the subsequent node. Further, the last internal node shown in the Tree view was 'sg' (specific gravity of serum) of people prone for kidney diseases. The 'sg' values varied among them from 1.007 to 1.015. Hence, the leaf nodes out of this range were shown as 'notCKD' and those within the range were shown as CKD.

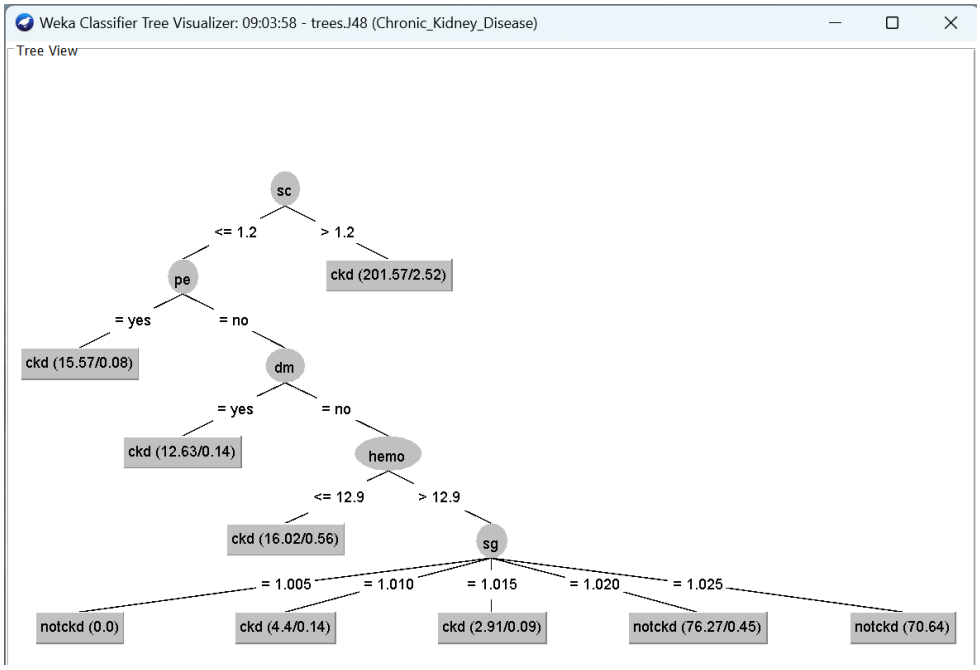

Figure 3. Tree view obtained from J48 classifier on the CKD dataset (UCI ML Repository) showing the risk disease markers successively tracing either CKD or NotCKD

The performance evaluation of the classifiers on CKD dataset was assessed under the heads Precision, Recall, F-Measure and Accuracy and the kappa stat revealed that they were all statistically significant as shown in Table 6. The least performance evaluation was obtained for the classifiers such as Stacking and Vote and found statistically insignificant.

While comparing the performance of each classifier from each folder in WEKA, the following best-performed classifiers were selected: Naïve Bayes, Multilayer Perceptron, Random Forest, Stacking and PART. The Multilayer Perceptron and Random Forest gave statistically significant performance ('v') compared to Naïve Bayes, Stacking and PART. The classifier Stacking performed insignificantly ('*') compared to others. While Naïve Bayes and PART performed better with the slightest difference from Multilayer Perceptron and Random Forest. The classifier Multilayer Perceptron yielded two MLP architecture (Figures 4 & 5) with input layer containing the chosen attributes, two hidden layers and an output layer showing the 'class'. As the relation between CKD attributes and the 'class' is non-linear, the hidden layers made intricate (Figure 4) and moderate (Figure 5) interconnections among the neurons showing the importance of the chosen attributes in assigning the 'class'.

Table 6. Performance evaluation by classifier

Classifiers	Precision (%)	Recall(%)	F-Measure (%)	Accuracy (%)	Kappa stat*
NaïveBayes	100	92	95.8	95.00	0.9164
Logistic	97.9	95.2	96.6	95.75	0.9102
MultilayerPerceptron	100	99.6	99.8	99.75	0.9947
SMO	100	96.4	98.2	98.25	0.9526
Voted Perceptron	62.5	100	76.9	97.5	0 (NS)
Bagging	99.6	98.4	99	98.75	0.9734
Logit Boost	100	99.6	99.8	99.25	0.9947
Stacking	62.5	100	76.9	62.5	0 (NS)
Vote	62.5	100	76.9	62.5	0 (NS)
Decision Table	98.8	99.6	99.2	98.75	0.9786
PART	98.8	98.8	98.8	98.5	0.9680
J48	98.8	99.6	99.2	99	0.9786
Random Forest	100	100	100	100	1.0000
Random Tree	97.9	94.8	96.3	96	0.905

*P<0.0001; NS: Not significant

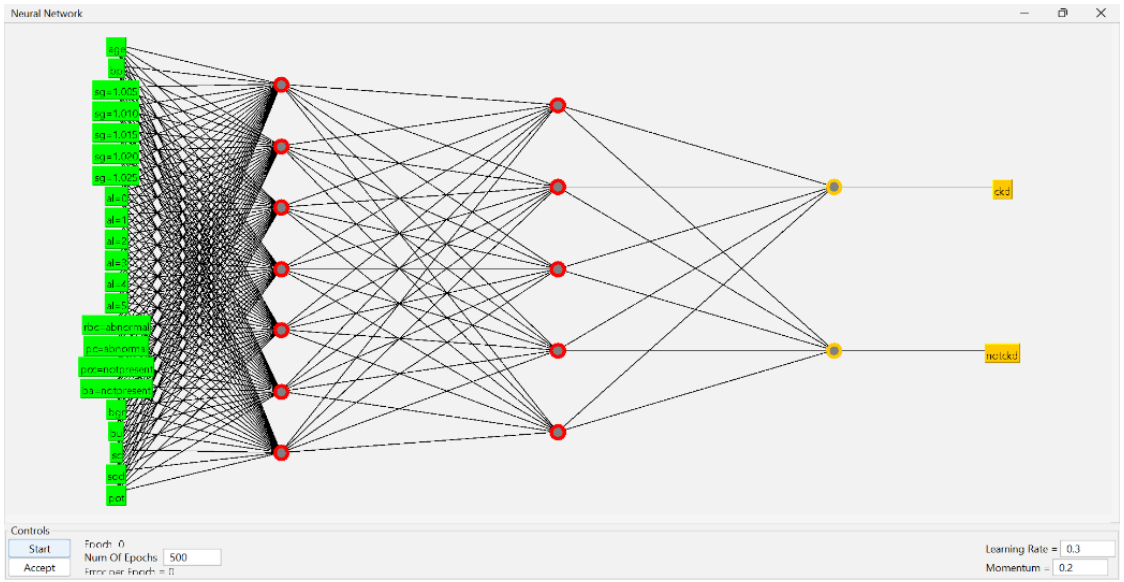

Figure 4. Structure of MLP architecture using the dataset attributes (age, bp, sg, al, rbc, pc, pcc, ba, bu, sc, sod and pot) in two hidden layers. The Feed-forward neural network connecting all the neurons appears to be intricately connected to the dataset attributes chosen.

The edited test dataset comprising 10 instances, with 'class' attribute kept as blank, chosen randomly out of the CKD dataset was allowed to test using Random Forest Classifier, which predicted with high accuracy as authenticated in the CKD dataset.

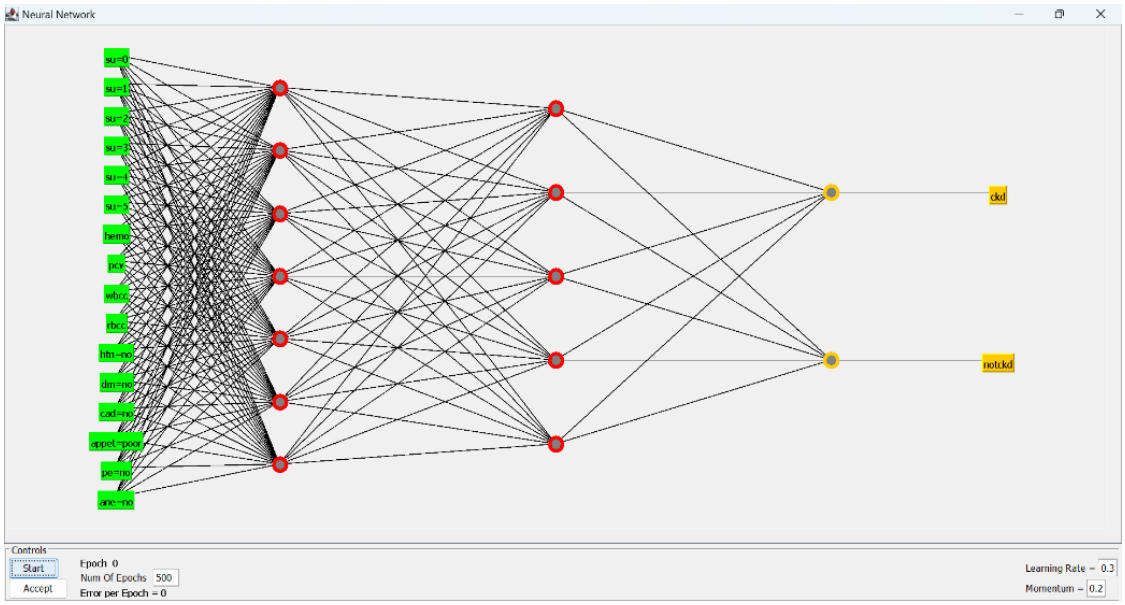

Figure 5. Structure of MLP architecture using the dataset attributes (su, hemo, pcv, wbcc, rbcc, htn, dm, cad, appet, pe and ane) in two hidden layers. The Feed-forward neural network connecting all the neurons appears to be less intricately connected to the dataset attributes chosen.

Discussion

The Random Forest classifier yielded a significant performance compared to the rest of the classifiers used in our study. Further, the decision tree (Figure 3) obtained using J48 classifier depicted the events showing the risk factors successively in building the severity of CKD. Random Forest, Multilayer Perceptron, Logit Boost, J48, Decision Table, Bagging, PART and SMO classifiers have yielded the kappa statistics with maximum statistically significant values between 0.9526-1.000 (Table 3), indicating the perfect agreement relating to these classifiers' accuracy prediction.

The mean absolute error (MAE) was in the order from the least to the highest: Multilayer Perceptron < Logit boost < J48 < SMO < PART <Random Forest < Naïve Bayes<Logistic<Random Tree<Bagging<Decision Table. The MAE indicates the variation between the forecasted and the actual value and the lower MAE score obtained for the first six aforementioned classifiers used in the present study possibly suggest them as the reliable algorithms to be used for CKD dataset as MAE score is an indication of the matching of the predicted value with an actual value. The stacking, Vote and Voted Perceptron classifiers had high mean absolute error values (0.4689, 0.4689 and 0.375), possibly indicating unreliable accuracy prediction. Accordingly, the Root Mean Squared Error (RMSE) values obtained reflected the standard deviation of residuals (Table 3) and the lower RMSE values shown by Multilayer Perceptron, J48 and Random Forest compared to other classifiers possibly qualify them as reliable for accuracy prediction using CKD dataset and followed the general trend that the greater the forecast and accuracy, the lower the RMSE and MAE values and higher the specificity values (Table 4). The least time taken to build the model was obtained by the classifiers namely Random Tree and Vote with no time lapse (0 sec). Furthermore, the classifiers *viz.*, Stacking, SMO and Naïve Bayes ran within 0.01 seconds. And then, the classifiers namely PART and Voted Preceptron took 0.03 seconds. J48 classifier took 0.04 seconds. The Logistic, Decision Table and Multilayer Perceptron had the maximum time to build the model *viz.*, 0.09, 0.13 and 1.7 seconds respectively.

However, Stacking, Vote, and Voted Perceptron classifiers classified the instances least at 62.5%, suggesting that the Random Forest can be considered as the best algorithm for accuracy prediction (Table 6).

The treeview (Figure 3) developed for the CKD dataset by the J48 classifier would be one of the best guides or decision trees useful for the suspected kidney patient to undertake the precautionary prophylactic measures.

Out of the 14 classifiers used in the evaluation of the CKD dataset, Random Forest yielded 100% Precision, Recall, F-Measure and Accuracy. Equally, well performance evaluation parameters were obtained for Multilayer Perceptron, Logit Boost and J48 (Table 6). The aforementioned classifiers were used in various applications such as EEG signals by Kokluand Sabanci [26] and Bharati et al. [27], Decision Tree model in non-communicable diseases [28] and Stacking and Bagging in the prediction of heart disease risk [29]. Haldar et al. [30] employed J48, Decision Tree, ZeroR, and Naïve Bayes algorithms on haematological data from individuals with diabetes and reported that Naïve Bayes with an accuracy of 76.30% was the best algorithm for diabetes data. The comparative evaluation of the J48, Decision tree, Multilayer Perceptron and Naïve Bayes was done by Amin and Habib [31] and demonstrated that J48 has an accuracy of 97.61% and proposed that it was the best algorithm and then Naïve Bayes, which has the lowest error rate of 27.91% was the second-best algorithm [31]. The compilation of different ML based tools used on MLKDD was categorized by Qezelbash-Chamak [12]. The feature optimization approaches with voting ensemble model have been reported as highly appropriate in the diagnosis of CKD [32]. However, in the present study, the Radom Forest yielded the highest accuracy with the CKD dataset retrieved from the UCI repository

Conclusion

The tree view developed for the CKD dataset using the J48 classifier is found to be the best guide or decision tree, which depicted synchronously and successively the role of each attribute (disease markers) in tracing the prevalence of CKD in a prospective patient and the same possibly helps to undertake the precautionary prophylactic measures. The Random Forest and Multilayer Perceptron, Logit Boost and J48 gave the best performance evaluation among the 14 classifiers used. Random Forest gave the highest accuracy, precision, recall and F-measure, hence, it is recommended to be used.

List of abbreviations

ARFF: Attribute Related File Format
CKD: Chronic Kidney Disease
CKDu: Chronic Kidney Disease of Unknown Etiology
eGFR: Estimated glomerular filtration rate
ESRD: End stage renal disease
J48: Algorithm to generate decision tree
ML: Machine Learning
MLP: Multilayer Perceptron
MLKDD: Machine learning Kidney Disease Diagnosis
RRT: Renal replacement therapy
SMO: Sequential Minimal Optimization
UCIMLR: University of California Irvine Machine Learning Repository
WEKA: Waikato Environment for Knowledge Analysis

Conflict of Interest

The authors declare that they have no conflict of interest.

Authors' Contributions

Gayathri Sakhamuri explored the dataset, identified UCI machine learning repository, shown

interest on CKD dataset, performed experiment with WEKA using 14 classifiers and wrote the Introduction and Methodology sections. Krupanidhi Sreerama guided in exploring WEKA software tool, explored WEKA tool in 'Experimenter' Environment, wrote the Results and Discussion sections. Both authors revised the manuscript.

Acknowledgements

The authors acknowledge the Head of the Department of Biotechnology, VFSTR University, India for providing Centre of Excellence facilities during the tenure of the present work.

References

1. Webster AC, Nagler EV, Morton RL, Masson P. Chronic kidney disease. Lancet 2017;389(10075):1238-1252.
2. Abraham G, Agarwal SK, Gowrishankar S, Vijayan M. Chronic Kidney Disease of Unknown Etiology: Hotspots in India and Other Asian Countries. Semin Nephrol. 2019;39(3):272-277. doi: 10.1016/j.semnephrol.2019.02.005
3. Ganguli A. Uddanam nephropathy/regional nephropathy in India: preliminary findings and a plea for further research. American Journal of Kidney Diseases 2016;68(3):344-348.
4. Suchitra M, Mahapatra R. Kidney conundrum. Down to earth. 2013 [Online]. Available from: http://www.downtoearth.org.in/coverage/kidney-conundrum-42845.
5. Kakitapalli Y, Ampolu J, Madasu SD, Kumar MS. Detailed review of chronic kidney disease.Kidney Diseases. 2020;6(2):85-91. doi:10.1159/000504622
6. Mascarenhas S, Mutnuri S, Ganguly A. Deleterious role of trace elements - silica and lead in the development of chronic kidney disease. Chemosphere. 2017;177:239-249.
7. Evans PD, Taal MW. Epidemiology and causes of chronic kidney disease. Medicine.2011; 9(7):402-406.
8. Coresh J, Selvin E, Stevens LA, Manzi J, Kusek JW, Eggers P, et al. Prevalence of chronic kidney disease in the United States. JAMA 2007;298:2038-2047.
9. Kump P M, Xia J, Yaddanapudi S, Bai E. An automated treatment plan alert system to safeguard cancer treatments in radiation therapy.Machine Learning with Applications2022;10:100437.doi: 10.1016/j.mlwa.2022.100437
10. Osman SMI, Sabit A. Predictors of COVID-19 vaccination rate in USA: A machine learning approach.Machine Learning with Applications. 2022;10:100408. doi:10.1016/j.mlwa.2022.100408
11. Jiang H, Salmon BP, Gale TJ, Dargaville PA. Prediction of bradycardia in preterm infants using artificial neural networks. Machine Learning with Applications2022;10:100426.
12. Qezelbash-Chamak J, Badamchizadeh S, Eshghi K, Asadi Y. A survey of machine learning in kidney disease diagnosis.Machine Learning with Applications. 2022;10:100418 doi:10.1016/j.mlwa.2022.100418
13. Zupan B, Halter JA, Bohanec M. Qualitative model approach to computer assisted reasoning in physiology. Proceedings of Intelligent Data Analysis in Medicine and Pharmacology-IDAMAP98. 1998.
14. Xiuyi T, Yuxia G. Research on Application of Machine Learning in Data Mining, IOP Conf. Series: Materials Science and Engineering. 2018;392(6):062202.
15. McClellan WM, Flanders WD. Risk factors for progressive chronic kidney disease. J Am Soc Nephrol. 2003; 14(suppl 2):S65–70. doi:10.1097/01.asn.0000070147.10399.9e
16. Dhillon A, Singh A. Machine learning in healthcare data analysis: a survey. Journal of Biology and Today's World. 2019;8(6):1-10.
17. Bai Q, Su C, Tang W, Li Y. Machine learning to predict end stage kidney disease in chronic kidney disease. Sci Rep. 2022;12(1):8377. doi: 10.1038/s41598-022-12316-z

18. Singh V, Asari VK, Rajasekaran R. A deep neural network for early detection and prediction of chronic kidney disease.Diagnostics. 2022;12(1):116.doi:10.3390/diagnostics12010116
19. Iftikhar H, Khan M, Khan Z, Khan F, Alshanbari HM, Ahmad ZA. Comparative Analysis of Machine Learning Models: A Case Study in Predicting Chronic Kidney Disease. Sustainability.2023;15(3):2754.doi:10.3390/su15032754
20. Ghassemi M, Naumann T, Schulam P, Beam AL, Chen IY, Ranganath R. A review of challenges and opportunities in machine learning for health. AMIA. 2020; Summits on Translational Science Proceedings 2020, p. 191.
21. Sevakula RK, Au-Yeung WTM, Singh JP, Heist EK, Isselbacher EM, Armoundas AA. State-of-the-art machine learning techniques aiming to improve patient outcomes pertaining to the cardiovascular system.Journal of the American Heart Association.2020;9(4):e013924. doi:10.1161/JAHA.119.013924
22. Park DJ, Park MW, Lee H, Kim YJ, KimY, Park YH. Development of machine learning model for diagnostic disease prediction based on laboratory tests.Scientific Reports. 2021;11(1):7567. doi:10.1038/s41598-021-87171-5
23. Wong LW, Tan GW, Ooi KB, Lin B, Dwivedi YK. Artificial intelligence-driven risk management for enhancing supply chain agility: A deep-learning-based dual-stage PLS-SEM-ANN analysis. International Journal of Production Research. 2022;20:1-21.
24. RubiniL, SoundarapandianP, EswaranP. 2015. Chronic_Kidney_Disease. UCI Machine Learning Repository. [Internet]. Available from: https://doi.org/10.24432/C5G020.
25. Merlini D, Rossini M. Text categorization with WEKA: A survey.Machine Learning with Applications. 2021;4:100033. doi:10.1016/j.mlwa.2021.100033.
26. Koklu M, Sabanci K. The classification of eye state by using kNN and MLP classification models according to the EEG signals.International Journal of Intelligent Systems and Applications in Engineering2015;3(4):127-130. doi:10.18201/ijisae.75836
27. Bharati S, Podder P, Raihan-Al-Masud M. EEG Eye State Prediction and Classification in order to Investigate Human Cognitive State 2018. 2018 International Conference on Advancement in Electrical and Electronic Engineering 2018;22:1-4.
28. Jeewandara N, Asanka PP. Data mining techniques in prevention and diagnosis of non communicable diseases. Int. J. Res. Comput. Appl. Robot. ISSN. 2017, pp. 2320-7345.
29. Latha C BC, Jeeva SC. Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques.Informatics in Medicine Unlocked2019;16:100203. doi: 10.1016/j.imu.2019.100203.
30. Haldar A, Raj GP, Lakshmi SV. Comparison of Different Classification Techniques Using WEKA for Diabetic Diagnosis. International Journal of Innovative Research in Computer and Communication Engineering. 2018;6(3):85-97.doi:10.4236/jsea.2013.63013
31. Amin N, Habib A. Comparison of Different Classification Techniques using WEKA for Hematological Data. American Journal of Engineering Research 2015;4(3):55-61.
32. Hossain MM, Swarna RA, Mostafiz R, Shaha P, Pinky LY, Rahman MM, et al. Analysis of the performance of feature optimization techniques for the diagnosis of machine learning-based chronic kidney disease. Machine Learning with Applications.2022;9:100330. doi:10.1016/j.mlwa.2022.100330