A Study of Machine Learning Algorithms Performance Analysis in Disease Classification

Jai Kumar B[1,*] and Mohanasundaram R[1]

[1]*School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India*

Abstract:

Background: Because there are no symptoms, it might be difficult to detect CKD in its early stages. One of the main causes of CKD is diabetes mellitus (DM), and early detection of the condition can assist individuals in obtaining prompt treatment. Because this illness has no early signs and is only discovered after the kidneys have gone through 25% damage, early-stage prediction is not very likely. This is the key factor driving the need for early CKD prediction.

Objective: The objective of the paper is to find the best-performing learning algorithms that can be used to predict chronic kidney disease (CKD) at an earlier stage.

Methods: This research aimed to compare different machine learning algorithms used in different disease predictions by various researchers. In this comparative study, machine learning algorithms like Logistic Regression, K-Nearest Neighbor, Decision Tree, Support Vector Machine, Artificial Neural Network, Random Forest, Composite Hypercube on Iterated Random Projection, Naïve Bayes, J48, Ensembling, Multi-Layer Perceptron, Deep Neural Network, Autoencoder, and Long Short-Term Memory are used in disease classification.

Results: Each classification model is well tested in a different dataset, and out of these models, RF, DNN, and NB classification techniques give better performance in Diabetes and CKD prediction.

Conclusion: The RF, DNN, and NB classification algorithms worked well and achieved 100% accuracy in predicting diseases.

Keywords: Diabetes mellitus (DM), Chronic kidney disease (CKD), Random forest (RF), Naïve bayes (NB), Deep neural network (DNN), WHO.

*Address correspondence to this author at the School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India;
E-mail: jaikumar.b2020a@vitstudent.ac.in

1. INTRODUCTION

1.1. Overview

Due to its high mortality rate, chronic kidney disease, also known as CKD, has recently received a lot of media attention. The World Health Assembly of the WHO claims that developing nations are in danger from chronic illnesses. Unbeknownst to many, it is a disorder that frequently stays undiagnosed until the sickness is well along. The identification of CKD is often made *via* a physical examination of individuals who have been identified to be in danger of acquiring renal problems [1]. If CKD is detected in its early stages, it may be treated, but if it is not, itresults in renal failure. According to a study done in 2016, chronic kidney disease or chronic renal illness claimed the lives of 753 million individuals globally. Four hundred seventeen million of the 753 million individuals were women, and the other 336 million

were men. This results from chronic renal disease, which affects the urine system. As the kidneys' ability to operate declines over time, they eventually fail. The waste product present in the blood also causes various health issues such as diabetes, cardiovascular disease, eye problems and blood pressure. Gloabally, nearly 10% of people were suffering from this chronic kidney disease [2]. In China, 10.8% were affected [3], and in the United States ranges from 10% to 15% were affected [4]. From another survey, 14.7% of people were affected in the Mexican population [5]. The total loss of renal function is the primary symptom of this CKD condition. In its early stages, this disease will not show any of the symptoms; hence, it is identified after we lose kidney function by 25% [6]. They exhibit acute malfunction in immunological and neural disorders, which have detrimental effects on patients who are affected in their daily routines by chronic kidney disease (CKD). Kidney failure is brought on by the major problem in CKD. It leads to kidney transplant or else it leads to end-stage life span [7]. Therefore, by providing appropriate therapy and being able to anticipate CKD, the effects may be minimized. Here are a number of diagnostics and computing methods utilized in both the detection and evaluation of CKD complexity [8]. People from developed countries were not aware of this disease, which led to dialysis or transplants. To measure CKD, we have a variety of approaches, including computational ones. By assessing the rate at which glomerular filtration occurs (GFR) based on the patient's age, blood type, sex, cholesterol levels, weight, and other characteristics, this condition was discovered. The GFR value is classified into five stages [9]. Table **1** shows different GFR levels.

Table 1. Glomerular filtration rate (GRF) levels.

Stages	GFR (mL/min/ 1.73 m2)	Description
I	≥90	Kidney Function Fair
II	60–89	Slight CKD
III	30–59	A Medium CKD
IV	15–29	Severe CKD
V	≤ 15	End Stage CKD

To measure the GRF, we need two test values: one is from the blood test to check creatinine, and another is the urine test to check albumin [10]. This traditional method is used to identify the presence of albumin and creatinine, which calculates the GFR and the levels of kidney functions.

One of the main causes of chronic kidney disease (CKD) is Diabetes Mellitus (DM), and those with the condition are more likely to develop renal failure. People affected by diabetes have a lifelong disorder; taking regular treatment and prevention is very difficult; the medical expenses are also high, and the price is not affordable for average and below-average people. Regular monitoring and giving treatment are also tough for doctors. Because the pancreas does not create enough insulin, this kind of diabetes develops. Diabetes is mainly classified into three categories: gestational diabetes (Diabetes during pregnancy), type 1 (T1DM), and type 2

diabetes (T2DM). Insulin secretion by the pancreas fails in type 1 diabetes [11], hence the need of taking insulin physically. Type 2 diabetes occurs when not enough insulin is secreted from the pancreas to filter proteins in the blood. Gestational diabetes occurs during pregnancy due to an excess intake of food that causes insufficient insulin to process the blood. In this modern culture, CKD affects most people around the globe, and Type 2 diabetes is also one of the root causes of CKD. Hence, reducing the disease, monitoring it, and finding it is most important here. Patients with nephropathy or heart problems might have increased GFR values; such patients missed out on this CKD criteria. Furthermore, there is not enough testing and monitoring equipment, as well as inadequate doctors to observe every patient with CKD every time. The computer-based monitoring system needed for CKD diagnosis and testing is economical and affordable. Artificial intelligence techniques are used in most areas of the field for early prediction [12]. In the medical field, for disease prediction and classification, an advanced technique called machine learning is very useful for earlier detection. The data preprocessing and classification are done using Machine learning.

The intention of investigating Chronic Kidney Disease, also referred to as CKD, is to find the medical condition at an early stage and stop it from developing into end-stage renal failure, which is severe in the absence of kidney transplantation or external filtration (kidney function). Being able to prevent a medical condition's development *via* prompt care makes the diagnosis of chronic renal failure (CKD) crucial. The likelihood of a pause in the development of the illness increases with early detection. A correct diagnosis also makes it possible to start taking medications and altering the way one lives in order to control the illness and stop the consequences. Furthermore, knowing the underlying factors underlying kidney failure may aid in creating plans aimed at preventing the illness from ever starting in its preliminary region.

1.2. Paper Contribution

The next few reasons encapsulate all primary motives behind this work.

- In order to choose the best characteristics with regards to machine learning, researchers investigate several methods for selecting features.
- We aim to identify the most suitable feature group that has been verified clinically and also chosen from a machine learning standpoint.
- Collaborating through a healthcare professional, we identify the highest impacted aspects through the clinician's viewpoint as we examine the clinical diagnosis of chronic kidney disease.
- Using several subgroups of features from the chronic kidney disease set of data, researchers develop and analyse base-learning models.
- Our objective is to evaluate each system's efficiency in comparison to comparable variants.

1.3. Literature Review

The state of the art of CKD evaluation and prediction is addressed in this section. Our major areas of interest are feature-selection techniques and Machine Learning models with Deep Learning. Most studies focus on previous and laboratory test records to build promising predictive models. For example, Gazi Mohammed Ifraz *et al*. [13] employed machine learning techniques to classify and predict chronic kidney disease from the CKD dataset available in Kaggle. From this approach, the Logistic Regression (LR) performs well and produces an accuracy of 97%. Another approach by Hashi *et al*. [14] for the presence and absence of CKD prediction was done through the K-Nearest Neighbor (KNN) classification algorithm with 76.96% accuracy. In the study from Alasker *et al*. [15], which uses a Decision Tree (DT) classifier for predicting kidney disease, this Decision Tree (DT) results in 98.4127% accuracy by using all 24 features from the dataset taken from the UCI ML repository. Baitharu and Pani [16] employed an artificial neural network (ANN) for classifying the disease in the healthcare system using a dataset of liver disease, which produced an accuracy of 71.59%. In their study, Bashir *et al*. [17] developed a disease prediction model by using a multilayer classifier. They used K-Nearest Neighbour (KNN) for early disease prediction, which achieved 57.41% accuracy from the Statlong dataset. Khan *et al*. [18] utilized a KNN and Random Forest (RF) classifier to predict liver disease from the dataset taken from the UCI ML repository, KNN produced an accuracy of 62.90% and RF of 72.17% accuracy. An approach from Vijayarani *et al*. [19] employed an ANN classification to predict kidney disease in a dataset taken from different health centres like hospitals, labs, and medical centres. This ANN classification achieves an accuracy of 87.70%. Another approach by Dar and Azmeen [20] uses this decision tree (DT) classification technique for predicting dengue fever based on real data collected from various hospitals. This DT classification produced an accuracy of 76%. An approach from Pahareeya *et al*. [21] proposed a classification model for liver disease prediction from the liver dataset in the UCI ML repository. This SVM achieves an accuracy of 71.5026%. Khan *et al*. [22] employed a Composite Hypercube on Integrated Random Projection (CHIRP) for classifying kidney disease as positive or negative, using the dataset taken from the UCI ML repository. This CHIRP produces an accuracy of 99.75%. Rashed-Al-Mahfuz *et al*. [23] use the Random Forest (RF) classifier to predict the CKD or NOTCKD from the CKD dataset, which was taken from the UCI ML repository. RF achieves an accuracy of 99.50% for the selected features. Jeong *et al*. [24] used the Autoencoder (AE) for classifying the chronic kidney disease stages with a highly imbalanced dataset taken from the National Health Insurance Corporation (NHIC) in Korea. This AE achieves an accuracy of 99.58%. Rahman *et al*. [25] developed a diabetes disease prediction classification model using Conv-LSTM from PIDD in the National Institute of Diabetes and Digestive Diseases. The Conv-LSTM-based technique performs well and achieves 97.26% accuracy.

Garca-Ordás *et al*. [26] proposed a new approach based on the combination of a Sparse Autoencoder and a Convolutional classifier in diabetes classification from the Pima Indians Diabetes Dataset (PIDD). This model achieves an accuracy of 92.31% in disease classification. Nadesh *et al*. [27] developed a model to predict Type 2 diabetes mellitus from feature selection and Deep Neural networks. This proposed model performs well and achieves 98.16% accuracy using the Pima Indians Diabetes Dataset (PIDD) taken from the UCI ML repository. Senan *et al*. [28] used various ML classification techniques for CKD disease prediction like Random Forest, K-Nearest Neighbors, Decision Tree, and Support Vector Machine. The CKD dataset of 400 patients was collected from the UCI ML repository. This approach gives the best diagnostic results in disease classification and produces accuracy ranges of SVM 96.67%, KNN 98.33%, DT 99.17, and RF 100%. Sunge *et al*. [29] developed a diabetes prediction model using different Machine learning techniques, including Decision Tree (DT), K-Nearest Neighbors (KNN), Naïve Bayes (NB), and Artificial Neural Network (ANN) datasets taken from the Pima Indians Diabetes Dataset (PIDD). From this model, the KNN model achieves an accuracy of 80.34% in diabetes prediction. Another approach by Ranjith *et al*. [30] proposed a model to predict diabetes disease using various ML algorithms. The algorithms are K-Nearest Neighbors (KNN), Decision Tree (DT), Naïve Bayes (NB), Deep Neural Network (DNN), RL, and LR. With this algorithm, DNN performs well and produces 100% accuracy. Meng *et al*. [31] developed various ML models to predict pre-diabetes. These classification algorithms (LR, DT, and ANNs) are used to analyse risk factors and predict pre-diabetes. DT performs well and produces 77.87%. Aljaaf *et al*. [32] used the MLP neural network (MLP) to predict early-stage CKD; this approach achieves 98.1% accuracy. Subasi *et al*. [33] proposed a CKD diagnosis model using various Machine Learning (ML) classification techniques. The SVM, MLP, C4.5 DT, KNN, and RF were used, and Random Forest outperformed them well and achieved 100%. Another approach by Boukenze *et al*. [34] used MLP to predict Chronic Kidney failure; this model produces 99.75% accuracy. Almansour *et al*. [35] employed Neural networks (NN) and Support Vector machines (SVM) to predict CKD disease; this approach performed well, and both produced 97.75% accuracy. Gunarathne *et al*. [36] used a Decision Tree (DT) to predict CDK presence; this classification model achieves 99.1% accuracy. Another study by Kunwar *et al*. [37] uses two Machine learning (ML) techniques: Naïve Bayes (NB) and Artificial Neural Network (ANN) for CKD disease classification datasets taken from the UCI ML repository. Naïve Bayes performs well and produces an accuracy of 100%. Avci *et al*. [38] developed a disease classification model to predict CKD presence and absence. This model uses J48, SVM, NB, and K-Star to diagnose the UCI ML dataset for disease prediction. In this model, the J48 classifier achieves 99%. The study from Aliberti *et al*. [39] used NAR and an LSTM model for blood glucose prediction; this LSTM produces 88.55% accuracy in the

not filtered training set and 99.73% accuracy in the filtered training set. The RT_CGM dataset was taken from SAFHS. Pradhan *et al*. [40] employed Artificial Neural Network (ANN) for diabetes prediction. This model predicts diabetes with an accuracy of 85.09% based on a dataset taken from the Pima Indian Diabetes Dataset (PIDD). Islam *et al*. [41] used various machine learning (ML) techniques, but the ensembling approach outperformed the others in predicting Type 2 diabetes. This model produced 95.94% accuracy in the disease classification model. The study by Dritsas *et al*. [42] developed a CKD prediction model using the Random Forest (RF) classification algorithm. This RF diagnosis the CKD dataset from the UCI ML repository to predict CKD with an accuracy of 99.2% [43]. This paper presents an adaptive interference removal framework (IRF) for video person re-identification (V-ReID) to improve accuracy [44]. A novel feature learning framework is proposed to capture significant information in spatial and temporal domains, building a discriminative and robust feature representation for each sequence [45]. This study investigated the associations of behavioral and health-related factors with chronic kidney disease (CKD) in Iranian patients. A hospital-based case-control study found that low birth weight, diabetes history, kidney disease history, and chemotherapy history are associated with CKD risk. The results highlighted the importance of collaborative monitoring of kidney function among patients with these conditions [46]. To solve the shortcomings in current picture paint techniques, a suggested restored image consider blends linguistic assumptions with a thorough resultant group. It comprises of a Natural language Priors Network, which is Deep Attention Residual Company, and Full-scale Remove Connection. The approach concentrates on stream features learning logical prior knowledge for deficient areas, and repairs defective regions *via* full-scale jump joins. Current state-of-the-art approaches are surpassed by the method [47]. This study suggests a lightweight, single-image super-resolution system that combines multi-level characteristics to solve typical issues such as slowly convergent images and blurry edges. The suggested approach works much better when factors are large, and it operates present methods by means of individual viewpoints and independent metrics [48]. The study suggests a two-stage picture inpainting network that is superior and depends on situational perception and similar networks. Research on public datasets show that the technique delivers an increased likelihood of a mending outcome [49]. The study suggests a stochastic adversarial network-based technique for efficient picture inpainting. The combination of these modules enhances the visual impact and quality of the images, surpassing current standards in the two types of assessments [50]. In order, the level of imagine inpainting increases. This research suggests a simple method that employs grouping combination along with the attention mechanism. Comparing experimental findings with similar light-weight strategies, infer less time and used resources are observed.

2. MATERIALS AND METHODS

In the following phase, we will deeply discuss various algorithms in Machine learning that are employed to segregate the diseases. We have a large number of algorithms for designing the classification model. Here, we will see some of the classification techniques which are used. The algorithms are Logistic Regression (LR), K-Nearest Neighbour (KNN), Decision Tree (DT), Support Vector Machine (SVM) Artificial Neural Network (ANN), Random Forest (RF), Composite Hypercube on Iterated Random Projection (CHIRP), Naïve Bayes (NB), J48, Ensembling, Multi-Layer Perceptron (MLP), Deep Neural Network (DNN), Autoencoder (AE), and Long Short-Term Memory (LSTM) used in disease classification.

2.1. Logistic Regression (LR)

Logistic regression is one of the accurate classification models used in the medical industry. For independent characteristics, the LR, which is usually used to forecast the class variable, produces a probability output of 0 or 1.

2.2. K-Nearest Neighbour (KNN)

K-Nearest Neighbour is the simplest method to classify the labels from the given dataset. The distance between the unlabelled instances in the class was measured with the nearest instance present in the class. It uses the k-Nearest Neighbours classifier.

2.3. Decision Tree (DT)

A decision tree is the most popular method used to classify the data; based on the decision, the tree will grow gradually. It is used for all kinds of real-time problems to make the right decision. DT has a higher number of layers, which reduces the performance and is also a little more complex. This has an overfitting problem in data handling.

2.4. Support Vector Machine (SVM)

Support Without using the probability approach, vector machines are a classification technique used to determine the class labels from the dataset. By using the same reasoning to get the class labels for linear equations. There will be two approaches to problem-solving: linearly separable and non-linearly separable.

2.5. Artificial Neural Network (ANN)

The fundamental function of neural networks is the use of neurons to carry out communication between the brain and other areas of the human body. Using artificial neural networks, we may employ this logic to convey information and carry out various actions.

2.6. Random Forest (RF)

A random forest is created by the combination of several Decision trees. The appropriate Subset is classified from the dataset through decision-making. The output from the several decision trees should be categorised based on the majority vote.

2.7. Composite Hypercube on Iterated Random Projection (CHIRP)

Using CHIRP, the prediction has the highest accuracy compared to normal prediction. For 2D prediction and symmetrical location prediction, it requires analysing a single set of data; the analytically effective methods are used by CHIRP. All predictions are grouped together by the CHIRP classifier and made into new data.

2.8. Naïve Bayes (NB)

Compared to numeric data, the Naive Bayes Classifier performs better, however it can only be used with a certain kind of dataset. One of the several categorization methods used in Bayesian learning is the naive Bayes algorithm. The Naive Bayes Classifier and Bayesian Belief Networks are two common algorithms that use Bayesian learning.

2.9. J48

Data collection in J48 is carried out *via* a divide-and-conquer top-down recursive strategy.

2.10. Ensembling

In Machine learning, ensembling is used to predict more accurately than a single classification model. This method will combine the different individual models for better performance in output prediction. Two methods were used here: stacking and voting. Soft approaches are used in voting in order to get a high probability, and the Staking procedure then yields the final forecast.

2.11. Multi-Layer Perceptron (MLP)

With a multi-layer perceptron, we may add additional layers to the model to forecast better outcomes. For greater performance in the additional layers, we apply a variety of activation functions. Step function, soft maximum function, sigmoid function, and more. All updated weights are to be done through the backpropagation approach in MLP.

2.12. Deep Neural Network (DNN)

Along with a Deep Neural Network (DNN), an artificial neural network, the input and output include numerous hidden models. This is the main characteristic of the Neural network: it processes complex progressive data in real-time. The forward propagation method is used for data processing from input to output.

2.13. Autoencoder (AE)

It is a data compression technique that is used for compressing the data in the hidden layer of a neural network. The encoder is used to decrease the hidden units in the input layer, and the decoder is used to increase the hidden units in the output layer. The error handling is done through the backpropagation method.

2.14. Long Short-Term Memory (LSTM)

Long Short-Term Memory is used to store and process data for a long time. To reduce the vanishing gradient problem, the LSTM was used. It has some memory space for temporary storage. LSTMs employ a Tanh layer and three logistic sigmoid gates to control memory. The output for this model is the binary numbers 0 or 1.

2.15. Proposed Method

With the purpose of comparing many cutting-edge machine learning techniques for CKD detection, approaches for selecting features served for which traits were best chosen. Medically, CKD is impacted by the selected characteristics. The chosen dataset, the preprocessing stage, the feature-selection techniques, the ML model, and the optimisation techniques comprise the whole suggested approach's stages. The outcomes might change based on each suggested machine learning model.

3. PERFORMANCE EVALUATION METRICS

3.1. Confusion Matrix

Table **2** shows the performance evaluation of the model.

Using the evaluation metrics, we evaluate the model to validate the performance. The metrics are Accuracy, Recall, Precision and F- measure.

Table 2. Confusion matrix.

Class	1(Positive)	0(Negative)
1(Positive)	TP-True Positive	FP-False Positive
0(Negative)	FN-False Negative	TN-True Negative

3.2. Accuracy

When more than half of a classification's predictions come true, it is said that the expected value and the actual value are equal.

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

3.3. Recall

Recall is used to calculate the system's ability to predict the number of positive samples

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

3.4. Precision

When a sample is projected to be positive, precision measures the proportion of actual samples that are positive.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3}$$

3.5. F-measure

A model's prediction accuracy is summed up by the F-measure, which is the harmonic mean of the precision and recall.

$$\text{F} - \text{Measure} = 2 \frac{(\text{Recall x Precision})}{(\text{Precision} + \text{Recall})} \tag{4}$$

where TP - True Positive, TN - True Negative, FP - False Positive and FN - False Negative, respectively.

4. RESULTS AND DISCUSSION

In this section we discuss the algorithm performance in the disease prediction from various aspects.

Table **3** represents the different algorithms used for disease prediction with different accuracy ranges in various diabetes datasets. This comparison shows that the DNN outperforms well in the diabetes dataset and produces higher accuracy than the remaining algorithms. Fig. (**1**) shows the pictorial representation of comparison results.

Table 3. Algorithm performs in diabetes dataset.

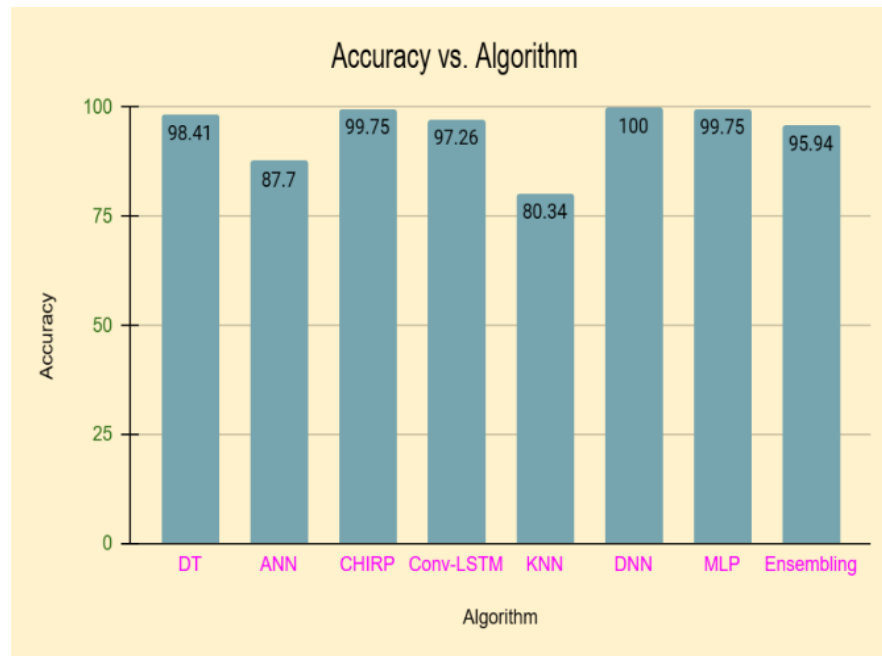| Author/Refs.	Algorithm	Accuracy (%)
Alasker *et al.* [15]	DT	98.41
Vijayarani *et al.* [19]	ANN	87.7
Khan *et al.* [22]	CHIRP	99.75
Rahman *et al.* [25]	Conv-LSTM	97.26
Sunge *et al.* [29]	KNN	80.34
Ranjith *et al.* [30]	**DNN**	**100**
Boukenze *et al.* [34]	MLP	99.75
Islam *et al.* [41]	Ensembling	95.94

Fig. (1). Algorithm comparison in the diabetes dataset.

Table **4** represents the different algorithms used for disease prediction with different accuracy ranges in various CKD datasets. This comparison shows that the RF and NB outperformed well in the CKD dataset and produced higher accuracy than the remaining algorithms. Fig. (**2**) shows the graphical representation of the results.

Table **5** represents the various algorithms used for disease prediction with different accuracy ranges in both the diabetes and CKD datasets. This comparison result shows that the DNN, RF, and NB outperformed well in both datasets and produced higher accuracy than the remaining algorithms. The graphical representation shown in Fig. (**3**).

Table 4. Algorithm performs in CKD dataset.

Author/Refs.	Algorithm	Accuracy (%)
Hashi *et al.* [14]	KNN	76.96
Jeong *et al.* [24]	AE	99.58
Aljaaf *et al.* [32]	MLP	98.1
Subasi *et al.* [33]	**RF**	**100**
Almansour *et al.* [35]	NN and SVM	97.75
Gunarathne *et al.* [36]	DT	99.1

(Table 6) contd.....

Author/Refs.	Algorithm	Accuracy (%)
Kunwar *et al.* [**37**]	**NB**	**100**
Avci *et al.* [38]	J48	99
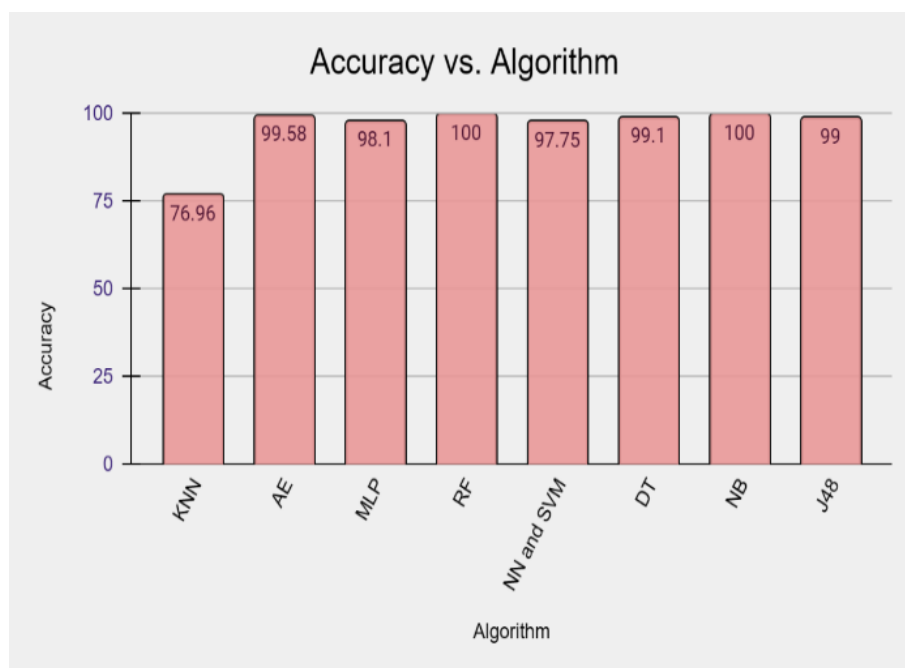
Fig. (2). Algorithm comparison in the CKD dataset.

Table 5. Algorithm performs in CKD and diabetes dataset.

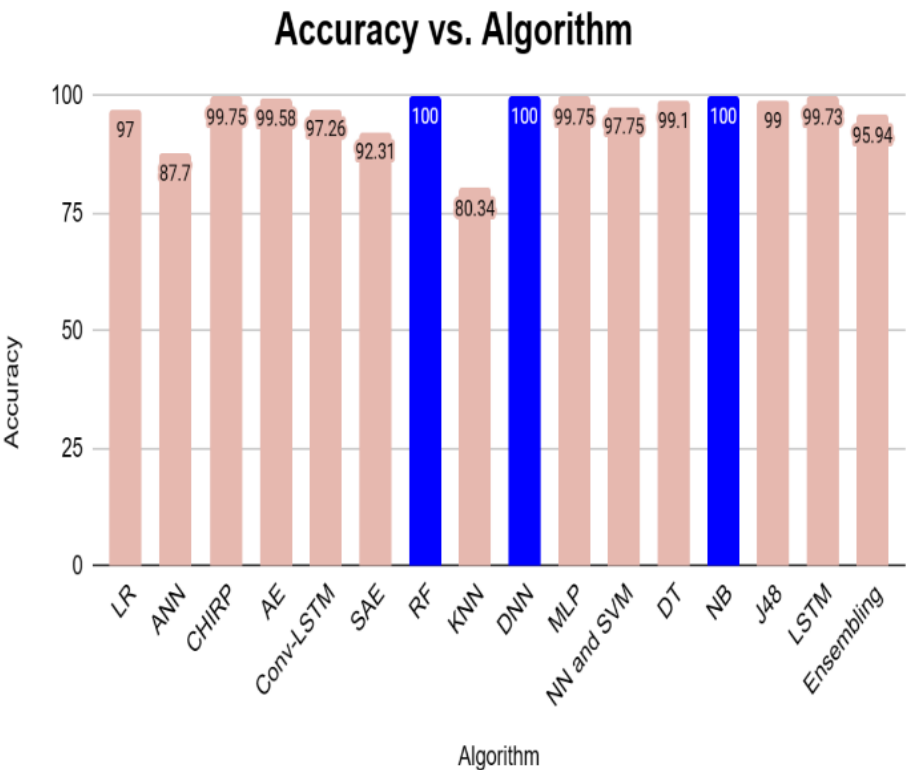| Algorithm	Accuracy (%)
LR	97
ANN	87.7
CHIRP	99.75
AE	99.58
Conv-LSTM	97.26
SAE	92.31
RF	100
KNN	80.34
DNN	100
MLP	99.75
NN and SVM	97.75
DT	99.1
NB	100
J48	99
LSTM	99.73
Ensembling	95.94

Fig. (3). Algorithm comparison in the diabetes and CKD dataset.

CONCLUSION

The comparative study observed that Machine Learning techniques will help to change the scenario in the medical sector. In this paper, we compared various algorithms for predicting the disease with different precision ranges from various studies. Finally, as observed from Table **5**, DNN, RF, and NB outperform well with an accuracy of 100%. This comparison study clearly highlights that DNN, RF, and NB are the most promising techniques for Disease prophecy.

Although the approach we recommend draws upon aspects that are shown to make a major impact on CKD classification, it makes sense from the standpoint of medicine. We want to make sure that our suggested methodology can be used broadly by expanding our research later on by obtaining real-world data in academic hospitals. In addition, since long-term kidney problems may coexist having various medical conditions, they are interested in exploring the relationship between the state of all medical conditions with the overall wellness of people having chronic renal failure. Finally, we want to investigate the produced technique's rate of computation.

LIST OF ABBREVIATIONS

DM	=	Diabetes Mellitus
CKD	=	Chronic Kidney Disease
RF	=	Random Forest
NB	=	Naïve Bayes
DNN	=	Deep Neural Network
WHO	=	World Health Organization

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

HUMAN AND ANIMAL RIGHTS

No humans/animals were used for studies that are the basis of this research.

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF DATA AND MATERIALS

The data supporting the findings of the article is available in the UCI ML repository, Kaggle and Pubmed at

URL
[https://archive.ics.uci.edu/dataset/336/chronic+kidney+disease,
https://archive.ics.uci.edu/dataset/857/risk+factor+prediction+of+chronic+kidney+disease,
https://pubmed.ncbi.nlm.nih.gov/22386035/,
https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database diabetes.csv].

CONFLICT OF INTEREST

To the knowledge of the authors, there are no conflicts of interest.

REFERENCES

[1] J. Qin, L. Chen, Y. Liu, C. Liu, C. Feng, and B. Chen, "A machine learning methodology for diagnosing chronic kidney disease", *IEEE Access,* vol. 8, pp. 20991-21002, 2020.
[http://dx.doi.org/10.1109/ACCESS.2019.2963053]

[2] Z. Chen, Z. Zhang, R. Zhu, Y. Xiang, and P.B. Harrington, "Diagnosis of patients with chronic kidney disease by using two fuzzy classifiers", *Chemom. Intell. Lab. Syst.,* vol. 153, pp. 140-145, 2016.
[http://dx.doi.org/10.1016/j.chemolab.2016.03.004]

[3] L. Zhang, F. Wang, L. Wang, W. Wang, B. Liu, J. Liu, M. Chen, Q. He, Y. Liao, X. Yu, N. Chen, J. Zhang, Z. Hu, F. Liu, D. Hong, L. Ma, H. Liu, X. Zhou, J. Chen, L. Pan, W. Chen, W. Wang, X. Li, and H. Wang, "Prevalence of chronic kidney disease in China: A cross-sectional survey", *Lancet,* vol. 379, no. 9818, pp. 815-822, 2012.
[http://dx.doi.org/10.1016/S0140-6736(12)60033-6] [PMID: 22386035]

[4] A. Singh, G. Nadkarni, O. Gottesman, S.B. Ellis, E.P. Bottinger, and J.V. Guttag, "Incorporating temporal EHR data in predictive models for risk stratification of renal function deterioration", *J. Biomed. Inform.,* vol. 53, pp. 220-228, 2015.
[http://dx.doi.org/10.1016/j.jbi.2014.11.005] [PMID: 25460205]

[5] A.M. Cueto-Manzano, L. Cortés-Sanabria, H.R. Martínez-Ramírez, E. Rojas-Campos, B. Gómez-Navarro, and M. Castillero-Manzano, "Prevalence of chronic kidney disease in an adult population", *Arch. Med. Res.,* vol. 45, no. 6, pp. 507-513, 2014.
[http://dx.doi.org/10.1016/j.arcmed.2014.06.007] [PMID: 24992221]

[6] H. Polat, H. Danaei Mehr, and A. Cetin, "Diagnosis of chronic kidney disease based on support vector machine by feature selection methods", *J. Med. Syst.,* vol. 41, no. 4, p. 55, 2017.
[http://dx.doi.org/10.1007/s10916-017-0703-x] [PMID: 28243816]

[7] S. Krishnamurthy, "Machine learning prediction models for chronic kidney disease using national health insurance claim data in Taiwan", In: *Healthcare.,* MDPI, 2021, p. 546.

[8] N. Bhaskar, and S. Manikandan, "A deep-learning-based system for automated sensing of chronic kidney disease", *IEEE Sens. Lett.,* vol. 3, no. 10, pp. 1-4, 2019.
[http://dx.doi.org/10.1109/LSENS.2019.2942145]

[9] S.M.M. Elkholy, A. Rezk, and A.A.E.F. Saleh, "Early prediction of chronic kidney disease using deep belief network", *IEEE Access,* vol. 9, pp. 135542-135549, 2021.
[http://dx.doi.org/10.1109/ACCESS.2021.3114306]

[10] R.Z. Alicic, M.T. Rooney, and K.R. Tuttle, "Diabetic kidney disease: Challenges, progress, and possibilities", *Clin. J. Am. Soc. Nephrol.,* vol. 12, no. 12, pp. 2032-2045, 2017.
[http://dx.doi.org/10.2215/CJN.11491116] [PMID: 28522654]

[11] T. Zhu, K. Li, P. Herrero, and P. Georgiou, "Basal glucose control in type 1 diabetes using deep reinforcement learning: An in silico validation", *IEEE J. Biomed. Health Inform.,* vol. 25, no. 4, pp. 1223-1232, 2021.
[http://dx.doi.org/10.1109/JBHI.2020.3014556] [PMID: 32755873]

[12] A. Sobrinho, A.C.M.D.S. Queiroz, L. Dias Da Silva, E. De Barros Costa, M. Eliete Pinheiro, and A. Perkusich, "Computer-aided diagnosis of chronic kidney disease in developing countries: A comparative analysis of machine learning techniques", *IEEE Access,* vol. 8, pp. 25407-25419, 2020.
[http://dx.doi.org/10.1109/ACCESS.2020.2971208]

[13] G.M. Ifraz, M.H. Rashid, T. Tazin, S. Bourouis, and M.M. Khan, "Comparative analysis for prediction of kidney disease using intelligent machine learning methods", In: *Computational and Mathematical Methods in Medicine,* vol. 2021. 2021.

[14] E.K. Hashi, M.S.U. Zaman, and M.R. Hasan, "An expert clinical decision support system to predict disease using classification techniques", *2017 International conference on electrical, computer and communication engineering (ECCE),* , 2017pp. 396-400
[http://dx.doi.org/10.1109/ECACE.2017.7912937]

[15] H. Alasker, S. Alharkan, W. Alharkan, A. Zaki, and L.S. Riza, "Detection of kidney disease using various intelligent classifiers", *2017 3rd international conference on science in information technology (ICSITech),* 2017pp. 681-684
[http://dx.doi.org/10.1109/ICSITech.2017.8257199]

[16] T.R. Baitharu, and S.K. Pani, "Analysis of data mining techniques for healthcare decision support system using liver disorder dataset", *Procedia Comput. Sci.,* vol. 85, pp. 862-870, 2016.
[http://dx.doi.org/10.1016/j.procs.2016.05.276]

[17] S. Bashir, U. Qamar, F.H. Khan, and L. Naseem, "HMV: A medical decision support framework using multi-layer classifiers for disease prediction", *J. Comput. Sci.,* vol. 13, pp. 10-25, 2016.
[http://dx.doi.org/10.1016/j.jocs.2016.01.001]

[18] B. Khan, R. Naseem, M. Ali, M. Arshad, and N. Jan, "Machine learning approaches for liver disease diagnosing", *Int. J. Data Sci. Anal.,* 2019.

[19] S. Vijayarani, S. Dhayanand, and M. Phil, "Kidney disease prediction using SVM and ANN algorithms", *Int. J. Comput. Bus. Res.,* vol. 6, no. 2, pp. 1-12, 2015.

[20] K. Shaukat Dar, S.M. Ulya Azmeen, S. Mehreen, and U. Azmeen, "Dengue fever prediction: A data mining problem", *J. Data Mining Genomics Proteomics,* vol. 6, no. 3, pp. 1-5, 2015.
[http://dx.doi.org/10.4172/2153-0602.1000181]

[21] J. Pahareeya, R. Vohra, J. Makhijani, and S. Patsariya, "Liver patient classification using intelligence techniques", *Int. J. Adv. Res. Comput. Sci. Softw. Eng.,* vol. 4, no. 2, pp. 295-299, 2014.

[22] B. Khan, R. Naseem, F. Muhammad, G. Abbas, and S. Kim, "An empirical evaluation of machine learning techniques for chronic kidney disease prophecy", *IEEE Access,* vol. 8, pp. 55012-55022, 2020.
[http://dx.doi.org/10.1109/ACCESS.2020.2981689]

[23] M. Rashed-Al-Mahfuz, A. Haque, A. Azad, S.A. Alyami, J.M.W. Quinn, and M.A. Moni, "Clinically applicable machine learning approaches to identify attributes of chronic kidney disease (CKD) for use in low-cost diagnostic screening", *IEEE J. Transl. Eng. Health Med.,* vol. 9, pp. 1-11, 2021.
[http://dx.doi.org/10.1109/JTEHM.2021.3073629] [PMID: 33948393]

[24] B. Jeong, H. Cho, J. Kim, S.K. Kwon, S. Hong, C. Lee, T. Kim, M.S. Park, S. Hong, and T.Y. Heo, "Comparison between statistical models and machine learning methods on classification for highly imbalanced multiclass kidney data", *Diagnostics,* vol. 10, no. 6, p. 415, 2020.
[http://dx.doi.org/10.3390/diagnostics10060415] [PMID: 32570782]

[25] M. Rahman, D. Islam, R.J. Mukti, and I. Saha, "A deep learning

approach based on convolutional LSTM for detecting diabetes", *Comput. Biol. Chem.,* vol. 88, p. 107329, 2020.
[http://dx.doi.org/10.1016/j.compbiolchem.2020.107329] [PMID: 32688009]

[26] M.T. García-Ordás, C. Benavides, J.A. Benítez-Andrades, H. Alaiz-Moretón, and I. García-Rodríguez, "Diabetes detection using deep learning techniques with oversampling and feature augmentation", *Comput. Methods Programs Biomed.,* vol. 202, p. 105968, 2021.
[http://dx.doi.org/10.1016/j.cmpb.2021.105968] [PMID: 33631638]

[27] B.M.K. P, S.P. R, N. R K, and A. K, "Type 2: Diabetes mellitus prediction using deep neural networks classifier", *Int. J. Cogn. Comput.,* vol. 1, pp. 55-61, 2020.
[http://dx.doi.org/10.1016/j.ijcce.2020.10.002]

[28] E. M. Senan, "Diagnosis of chronic kidney disease using effective classification algorithms and recursive feature elimination techniques", *J. Healthc. Eng.,* vol. 2021, 2021.
[http://dx.doi.org/10.1155/2021/1004767]

[29] A.S. Sunge, "Comparison data mining techniques to prediction diabetes mellitus", *J. Sustain. Eng.,* vol. 1, no. 2, pp. 225-230, 2019.
[http://dx.doi.org/10.35793/joseps.v1i2.31]

[30] M. Ranjith, H. Santhosh, and M. Swamy, "Machine learning algorithms for the detection of diabetes", *Int Res J Eng Technol.,* vol. 8, no. 01, pp. 135-140, 2021.

[31] X.H. Meng, Y.X. Huang, D.P. Rao, Q. Zhang, and Q. Liu, "Comparison of three data mining models for predicting diabetes or prediabetes by risk factors", *Kaohsiung J. Med. Sci.,* vol. 29, no. 2, pp. 93-99, 2013.
[http://dx.doi.org/10.1016/j.kjms.2012.08.016] [PMID: 23347811]

[32] A.J. Aljaaf, "Early prediction of chronic kidney disease using machine learning supported by predictive analytics", In: *2018 IEEE congress on evolutionary computation (CEC),* 2018, pp. 1-9.

[33] A. Subasi, E. Alickovic, and J. Kevric, "Diagnosis of chronic kidney disease by using random forest", *CMBEBIH 2017: Proceedings of the International Conference on Medical and Biological Engineering 2017,* , 2017pp. 589-594
[http://dx.doi.org/10.1007/978-981-10-4166-2_89]

[34] B. Boukenze, A. Haqiq, and H. Mousannif, "Predicting chronic kidney failure disease using data mining techniques", *Advances in Ubiquitous Networking 2: Proceedings of the UNet'16 2,* , 2017pp. 701-712
[http://dx.doi.org/10.1007/978-981-10-1627-1_55]

[35] N.A. Almansour, H.F. Syed, N.R. Khayat, R.K. Altheeb, R.E. Juri, J. Alhiyafi, S. Alrashed, and S.O. Olatunji, "Neural network and support vector machine for the prediction of chronic kidney disease: A comparative study", *Comput. Biol. Med.,* vol. 109, pp. 101-111, 2019.
[http://dx.doi.org/10.1016/j.compbiomed.2019.04.017] [PMID: 31054385]

[36] W. Gunarathne, K. Perera, and K. Kahandawaarachchi, "Performance evaluation on machine learning classification techniques for disease classification and forecasting through data analytics for chronic kidney disease (CKD)", *2017 IEEE 17th international conference on bioinformatics and bioengineering (BIBE),* 2017pp. 291-296
[http://dx.doi.org/10.1109/BIBE.2017.00-39]

[37] V. Kunwar, K. Chandel, A.S. Sabitha, and A. Bansal, "Chronic

kidney disease analysis using data mining classification techniques", *2016 6th International Conference-Cloud System and Big Data Engineering (Confluence),* 2016pp. 300-305
[http://dx.doi.org/10.1109/CONFLUENCE.2016.7508132]

[38] E. Avci, S. Karakus, O. Ozmen, and D. Avci, "Performance comparison of some classifiers on chronic kidney disease data", *2018 6th international symposium on digital forensic and security (ISDFS),* 2018pp. 1-4
[http://dx.doi.org/10.1109/ISDFS.2018.8355392]

[39] A. Aliberti, I. Pupillo, S. Terna, E. Macii, S. Di Cataldo, E. Patti, and A. Acquaviva, "A multi-patient data-driven approach to blood glucose prediction", *IEEE Access,* vol. 7, pp. 69311-69325, 2019.
[http://dx.doi.org/10.1109/ACCESS.2019.2919184]

[40] N. Pradhan, G. Rani, V.S. Dhaka, and R.C. Poonia, "Diabetes prediction using artificial neural network", In: *Deep Learning Techniques for Biomedical and Health Informatics.,* Elsevier, 2020, pp. 327-339.

[41] M.S. Islam, M.K. Qaraqe, S.B. Belhaouari, and M.A. Abdul-Ghani, "Advanced techniques for predicting the future progression of type 2 diabetes", *IEEE Access,* vol. 8, pp. 120537-120547, 2020.
[http://dx.doi.org/10.1109/ACCESS.2020.3005540]

[42] E. Dritsas, and M. Trigka, "Machine learning techniques for chronic kidney disease risk prediction", *Big Data Cogn. Comput.,* vol. 6, no. 3, p. 98, 2022.
[http://dx.doi.org/10.3390/bdcc6030098]

[43] H. Tao, Q. Duan, and J. An, "An adaptive interference removal framework for video person re-identification", *IEEE Trans. Circ. Syst. Video Tech.,* vol. 33, no. 9, pp. 5148-5159, 2023.
[http://dx.doi.org/10.1109/TCSVT.2023.3250464]

[44] W. Song, J. Zheng, Y. Wu, C. Chen, and F. Liu, "Discriminative feature extraction for video person re-identification via multi-task network", *Appl. Intell.,* vol. 51, no. 2, pp. 788-803, 2021.
[http://dx.doi.org/10.1007/s10489-020-01844-8]

[45] M. Ghelichi-Ghojogh, M. Fararouei, M. Seif, and M. Pakfetrat, "Chronic kidney disease and its health-related factors: A case-control study", *BMC Nephrol.,* vol. 23, no. 1, p. 24, 2022.
[http://dx.doi.org/10.1186/s12882-021-02655-w] [PMID: 35012483]

[46] Y. Chen, R. Xia, K. Yang, and K. Zou, "DARGS: Image inpainting algorithm via deep attention residuals group and semantics", *J. King Saud Univ. - Comput. Inf. Sci.,* vol. 35, no. 6, p. 101567, 2023.
[http://dx.doi.org/10.1016/j.jksuci.2023.101567]

[47] Y. Chen, R. Xia, K. Yang, and K. Zou, "MFFN: image super-resolution via multi-level features fusion network", *Vis. Comput.,* pp. 1-16, 2023.
[http://dx.doi.org/10.1007/s00371-023-02795-0]

[48] Y. Chen, R. Xia, K. Yang, and K. Zou, "DGCA: High resolution image inpainting via DR-GAN and contextual attention", *Multimedia Tools Appl.,* vol. 82, no. 30, pp. 47751-47771, 2023.
[http://dx.doi.org/10.1007/s11042-023-15313-0]

[49] Y. Chen, R. Xia, K. Zou, and K. Yang, "RNON: image inpainting via repair network and optimization network", *Int. J. Mach. Learn. Cybern.,* vol. 14, no. 9, pp. 2945-2961, 2023.
[http://dx.doi.org/10.1007/s13042-023-01811-y] [PMID: 37360881]

[50] Y. Chen, R. Xia, K. Yang, and K. Zou, "GCAM: lightweight image inpainting via group convolution and attention mechanism", *Int. J. Mach. Learn. Cybern.,* pp. 1-11, 2023.
[http://dx.doi.org/10.1007/s13042-023-01999-z] [PMID: 37360881]