Research Article

International Journal of Advanced Technology and Engineering Exploration, Vol 11(110)
ISSN (Print): 2394-5443 ISSN (Online): 2394-7454
http://dx.doi.org/10.19101/IJATEE.2023.10102081

Implementation of clinical diagnosis system for chronic kidney disease using deep learning algorithms

Ashwan A. Abdulmunem[1*] and Alaa Jamal Jabbar[2]
Department of Computer Science, Faculty of Computer Science and Information Technology, University of Kerbala, Kerbala, Iraq[1]
Directorate General of Dhi Qar Education, Shatrah Education Department, Dhi Qar, Iraq[2]

Received: 06-August-2023; Revised: 28-December-2023; Accepted: 31-December-2023

Abstract
Chronic kidney disease (CKD) ranks among the top 20 causes of death worldwide, affecting approximately 10% of adults. CKD impairs normal kidney function. The rising incidence of CKD underscores the need for effective prophylactic measures and early diagnosis. A novel aspect of this research is the development of a technique for diagnosing chronic renal diseases. This work aids researchers in exploring early detection methods for CKD prevention using deep learning (DL) techniques. The study involved the construction of a model for CKD diagnosis based on deep neural networks (DNN). A dataset of 400 patients with 24 features was analyzed, with mean and mode statistical analysis techniques employed to substitute missing numerical and nominal values. The effectiveness of the DNN has been demonstrated in the diagnosis results, achieving an accuracy of 98.33%. DL models employ complex algorithms to analyze large datasets containing various patient information, such as age, gender, lifestyle habits, and medical history. By automatically analyzing these factors together, the model can identify patterns indicative of potential kidney issues earlier than traditional methods. In addition to more accurate CKD prediction, DL models provide faster results, potentially leading to earlier interventions or treatments by physicians.

Keywords
Chronic kidney disease, Machine learning and Deep learning, Classification, Clinical diagnosis system.

1.Introduction
Many applications of machine and deep learning (DL) have been developed for the healthcare industry, including the identification of renal illness, diabetes, brain tumors, breast cancer, and cataracts. Millions of individuals worldwide suffer from kidney disease, which is the most prevalent healthcare consequence and greatly increases the chance of dying young [1]. Chronic kidney disease (CKD) is a non-communicable illness that has raised global patient admission rates, morbidity, and mortality rates [2]. It is spreading swiftly and rising to prominence as one of the leading causes of mortality globally [3]. According to a survey from 1990 to 2013, CKD is now the 13th biggest cause of mortality worldwide, with an annual rise in life loss of 90%. Kidney disorders are expected to affect 850 million individuals globally due to various factors [4].

A 2019 World Kidney day research states that kidney-related diseases claim the lives of at least 2.4 million people annually [5]. With an increasing prevalence worldwide, it is currently the sixth fastest-growing cause of death and is becoming into a serious public health issue. As detection, prevention, and treatment rates are still low in low-income nations, its impact is even greater [6].

Kidney disease is a serious health issue because of its ubiquitous pathophysiology, which is impacted by obesity, age, diabetes, and depression. These disorders have the potential to increase in prevalence [7]. According to the 2015 global burden of injuries, diseases, and risk factors assessment, the renal disease affect is 750 million people globally when the kidneys can no longer function normally, renal disease develops [8]. The kidneys can no longer remove waste and extra water from circulation, leading to a variety of health issues [9]. Renal activity is known to rise throughout months or years in patients with CKD. Furthermore, it has a

*Author for correspondence

decreasing glomerular filtration rate (GFR) of less than 60 m per min per 1.73m2. Chronic illnesses are the main cause of mortality and morbidity in India [10]. CKD are responsible for 60% of all mortality globally. In several nations, chronic illnesses resulted in 80% of deaths [11].

The National Renal Association categorizes CKD into stages depending on the occurrence of renal disorders and GFR, which is a marker of kidney function [12]. To lower the disease's high mortality rate, research should be widened and focused on the disease's early stages, diagnosing its risk group by laboratory testing, and preventing patients from progressing to the latter stages, such as dialysis, transplantation, or death [13]. A purpose for automated learning is to discover a substantial contribution so that early illness categorization may be carried out utilizing clinical laboratory findings [14]. Using machine learning's great potential in data analysis and classification [15]. It is critical that data-driven technical assistance tools enable rapid, precise, and cost-effective decision-making in the first diagnosis. They shorten the time necessary for diagnosis, allowing the patient to get therapy before the condition worsens beyond repair [16]. Machine learning is classified into three types: supervised learning, unsupervised learning, and reinforcement learning [17]. In order to deliver progressive and intelligent healthcare services, the present healthcare system is built utilizing cutting-edge competency technologies including big data analytics, machine learning, and information technology. Patients suffer globally as a result of the inadequate processing, filtering, and proper treatment of healthcare data, among other problems with the current healthcare system [18]. The fact that CKD is unpredictable due to its lack of dependence on a particular trait is one of the factors making diagnosis challenging. Furthermore, frequent CKD symptoms do not significantly aid in the diagnosis of the condition. The challenge of all previous researches was that it had been tested on small data sets [19–22].

The contribution of this work intends to create and execute DL model that utilizing clinical laboratory data, can predict the likelihood of CKD diagnosis in its early stages, cutting mortality and healthcare costs. The main objectives of the research are to design an automatic disease diagnosis system based on the latest DL techniques and the use of light weight modeling, which is represented by relying on the use of one dimension to reduce the time and complexity found in using two-dimensional networks.

The remainder of the paper is structured as follows: Section 2 presents related work on machine learning approaches in the field of CKD. Section 3 describes the CKD dataset and the proposed methods for CKD diagnosis. Section 4 presents the results and compares them with those of other researchers. The performance evaluation metrics and experimental results are discussed in Section 5. Finally, the conclusion and future work are illustrated in Section 6.

2.Literature review

The most prevalent type of machine learning applied in medical research is supervised learning [23]. There is an input object for each supervised learning instance (a comparator is commonly used, as the intended output value, sometimes referred to as a supervised signal) [24]. Techniques for supervised learning approaches that are often utilized include decision trees (DTs), naive bayes (NB) classification, least squares regression, logistic regression (LR), and support vector machine (SVM) algorithms[25–27]. Deep neural networks (DNN) have achieved expert-level performance in natural and biomedical image classification tasks, according to recent research [28] this, together with the capacity to develop hypotheses. Because of its flexibility to various data set processing and readily available open-source DL tools, DL plays an important role in encouraging medical research [29].

As a result, numerous efforts have been proposed to address this difficulty. One of these studies used deep learning application (DLA) based on retinal images to diagnose of CKD, with the goal of determining whether this method may be used as an adjunctive or opportunistic screening tool in appropriate situations. A DLA based only on retinal photos is comparable to material from a standard random forest (RF) model for CKD diagnosis, and it supports the possibility of employing retinal photography to diagnose CKD in some circumstances [30].

Salekin and Stankovic [31] on a dataset of 400 observations, evaluated classifiers like k-nearest neighbors (K-NN), RF, and artificial neural network (ANN). Five features were chosen for the study's model creation after the implementation of wrapper feature selection with the highest classification accuracy of 98% and a root mean square error (RMSE) of 0.11. A study by Debal and Sitote [32]

DT and SVM were employed to analyze a dataset with 400 occurrences and 14 parameters that achieved 96.75% accuracy rate. SVM is considered as a superior model.

Xiao et al. [33] used some of techniques such as LR, elastic net, least absolute shrinkage and selection operator (LASSO), ridge regression (RR), SVM, RF, XGBoost, neural networks, and K-NN. The authors proposed and compared models for forecasting the evolution of CKD. They discovered that when 551 patients' history data with proteinuria and 18 features were utilized, LR performed better with an accuracy of 0.873 in addition sensitivity and specificity of 0.83 and 0.82, respectively, and the result was classified as moderate, severe.

Islam et al. [34] used machine learning strategies to provide an early diagnosis of CKD have been investigated during this research. This subject has been the subject of a sizable amount of research. Nevertheless, this technique may be strengthened by utilizing predictive modeling. Investigate the relationship between data components and the traits of the target class in this technique to learn more. Beginning with 25 factors in addition to the class attribute, this study eventually reduced the list to 3 of those parameters, which it found to be the most effective subset for identifying CKD. A supervised learning environment has actually been used to investigate 12 different machine learning-based classifiers, with the boost classifier showing the best performance metrics with accuracy, precision, recall, and F1-Score values of 0.983, 0.98, 0.98, and 0.98, respectively.

Vinay et al. [35] using a hybrid RF classifier and two dimensional (2D) ultrasound kidney pictures, this work suggests an automated method for detecting CKD based on the textural characteristics of a kidney. Through testing on a dataset of 150 photos, the suggested classifier is compared to the other rival machine learning classifiers and provides a higher accuracy of 96.67%.

Qadir and Abd [36] this study uses a dataset of 12,446 computational thinking (CT) urogram and whole abdomen images to focus on kidney stones, cysts, and tumors-the three most common types of renal illness. The goal is to advance the field of artificial intelligence research while developing a kidney disease diagnosis system that is artificial intelligent AI-based. For the purpose of kidney illness picture identification, this work employs a hybrid technique that makes use of both pre-train models for feature extraction and classification using machine learning algorithms. The Densenet-201 model is the pre-trained model utilized in this investigation. With an accuracy rate of 99.719 percent, the Densenet-201-Random-Forest technique has surpassed many of the earlier models employed in other studies in addition to employing RF for classification.

Venkatrao and Kareemulla [37] suggested the deep separable convolution neural network (DSCNN), DL-based method, for the early identification of CKD. The capsule network (Caps Net) extracts more processing qualities of characteristics selected to signal a kidney problem. To expedite the categorization procedure, the relevant criteria are chosen using the aquila optimization algorithm (AO) approach. The required features require less computational work and increase classification effectiveness. Using sooty tern optimization algorithm (STOA), renal sickness is diagnosed as either CKD or non-CKD in DSCNN. The dataset accuracy is then tested using the CKD dataset, which can be obtained in the university of California Irvine (UCI) machine learning repository.

Alikhan et al. [38] this paper proposes the use of internet of things (IoT) and cloud computing in a smart medical big data health care system to diagnose CKD using a self-attention convolutional neural network optimized with season optimization algorithm (SACNN-SOA-CKD-IoT-CC). IOT devices, such as wearables and sensors, collect data. Applications of self-attention convolutional neural network (SACNN) include the diagnosis model for CKD. However, the SACNN does not disclose the use of any optimization algorithms to determine the ideal parameters and ensure that CKD is accurately classified. In comparison to the current approaches, the suggested SACNN-SOA-CKD-IoT-CC method obtains lower error rates of 11.27%, 8.35%, and 21.06%, as well as higher accuracy of 15.66%, 21.65%, and 9.64% in the CKD dataset.

Singh and Jain [39] for finding the most informative parameters for CKD diagnosis, the suggested method optimizes the SVM classifier using the hybridized dimensionality reduction strategy. It uses two phases to manage feature selection. The first uses the relief technique as a filter to give each feature in the dataset a weight and rank. The second phase involves using principal component analysis (PCA), a feature extraction method, to reduce the dimensionality of the best-selected subset. Multiple processors are used for simultaneous execution to process datasets more quickly. Using the clinical CKD dataset, the suggested model produced the best prediction accuracy, coming in at 92.5%.

Liang et al. [40] based on demographics, clinical, and comorbidity data, 8 machine learning models were employed to predict a patient with CKD will develop to end stage renal disease (ESRD) within three years. The most important markers were identified using LASSO, RF, and XGBoost. In addition, four sophisticated attribution methods were included to the DL model to improve model intelligibility. The accuracy of the DL model was 0.8991, which was substantially higher than that of the baseline models. DL using attribution techniques, RF and XGBoost interpretation was congruent with clinical knowledge, however LASSO-based interpretation was inconsistent. Hematuria, proteinuria, potassium, and the urine albumin to creatinine ratio were all linked with CKD development, although eGFR and urine creatinine were not.

Kaur et al. [41] this study used machine learning methods to predict renal disease using secondary data. Analysis methods include SVM, RF, LR, and K-NN, also present an ensemble machine learning method that combines RF, support vector clustering (SVC) and LR. This approach outperforms other classifiers in terms of accuracy. The algorithms' performance is measured using precision, recall, F1-Score, accuracy, and cohen kappa, outperforms other classifiers with 99% accuracy, 99% precision, 99% recall, 99% F1-Score, and 98% cohen kappa score.

Busi and Stephen [42] to explore the potential of several machine learning techniques, this study suggests an efficient extreme gradient boosting method for the rapid identification of kidney illness. Vector features are among the many properties that dense net is able to extract. Following the feature extraction stage, the data are passed to the feature selection stage. To choose the features, the improved salp swarm algorithm (ISSA) is employed. The recommended CKD classification technique was modeled using python. The CKD dataset from the UCI machine learning resources is then used to assess the dataset. Sensitivity, accuracy, and specificity serve as the suggested CKD classification technique's performance metrics. The results of the experiment indicate that the proposed strategy performs better for diagnosing CKD than the state-of-the-art approach at the moment.

Lu, et al. [43] this study includes data from 1,358 patients with CKD who were pathologically diagnosed at Zhongshan Hospital between December 2017 and September 2020. A machine learning-based CKD prediction interpretation framework was proposed. Using a recursive feature reduction with LR feature screening, 17 variables were selected from a set of 100 for model creation. Several machine learning classifiers were trained to predict 24-hour urine protein, including extreme gradient boosting, Gaussian-based NB, a neural network, RR, linear model LR, and an ensemble model. A global interpretation was used to evaluate the detailed relationship between the risk of developing CKD development and these variables. A patient-specific analysis was conducted with the assistance of these classifiers. The findings indicated that LR performed the best in a single machine learning model, with an area under the curve (AUC) of 0.850. The AUC was increased to 0.856 utilizing the ensemble model built using the vote integration approach.

Almansour et al. [44] this study employed ANN and SVM as classification methodologies. The mean of the related characteristics was used to replace all missing values in the dataset in order to conduct tests. Subsequently, by adjusting the parameters and conducting several trials, the ideal parameters for ANN and SVM approaches were found. The optimally acquired parameters and characteristics were utilized in the development of the final models for each of the suggested strategies. With accuracies of 99.75% and 97.75%, respectively, the actual findings of the studies showed that ANN outperformed SVM, suggesting that the conclusion of this study is quite promising.

Maisha et al. [45] this effort aims to provide a more dependable method that makes processing noisy data possible. However, the noisy and inconsistent results in the CKD dataset make it difficult to forecast CKD using conventional machine learning techniques. Therefore, using a mix of DNN, statistical techniques, PCA, and synthetic minority oversampling technique (SMOTE) this research proposes a way of handling noisy and erroneous values of the CKD dataset. As a result, a variety of machine learning techniques employ the improved CKD dataset that results from the previously discussed pre-processed techniques. Based on our findings, RF performed better than SVM, K-NN, DT, NB and LR classifiers, with an accuracy rate of 98.5%. Furthermore, research discovered that characteristics showing their dominance in outcome prediction were blood urea (BU) and serum creatinine (SC).

3.Methodology

This section will describe the dataset, pre-processing, and the proposed method based on using DL models to diagnose CKD in patients to depict the effectiveness of artificial intelligence algorithms in the medical imaging sector. The steps of the suggested method are shown in *Figure 1*.

The dataset for CKD is utilized in this research. This dataset has also been utilized by several researchers [8]. This dataset is made available on the UCI machine learning repository UCI. There are 400 occurrences in this dataset with 24 characteristics, including one target attribute. The target property is tagged with two classes to indicate CKD or not CKD. The dataset was compiled in 2015 from some hospitals. Additionally, it has a missing value. In *Table 1* below, each of the 24 qualities is described [9]. The dataset is available on Kaggle.

3.1Dataset

Table 1 The attributes present in the CKD dataset

Attributes	Type	Attributes	Type
Age	Numeric	Sodium	numeric
Blood pressure(BP)	Numeric	Potassium	numeric
Sugar	Nominal	Haemoglobin	numeric
Albumin	Nominal	Packet Cell Volume	numeric
Specific gravity(SG)	Nominal	White Blood Cell Count	numeric
Red blood cells(RBC)	Nominal	Red Blood Cell Count	numeric
Pus cell(PC)	Nominal	Hypertension	nominal(yes, no)
Pus cell clumps(PCC)	Nominal	Diabetes Mellitus	nominal(yes, no)
Bacteria	Nominal	Coronary Artery Disease	nominal(yes, no)
Blood glucose(BG)	Numeric	Appetite	nominal(good, poor)
BU	Numeric	Pedal Edema	nominal(yes, no)
SC	Numeric	Anemia	nominal(yes, no)

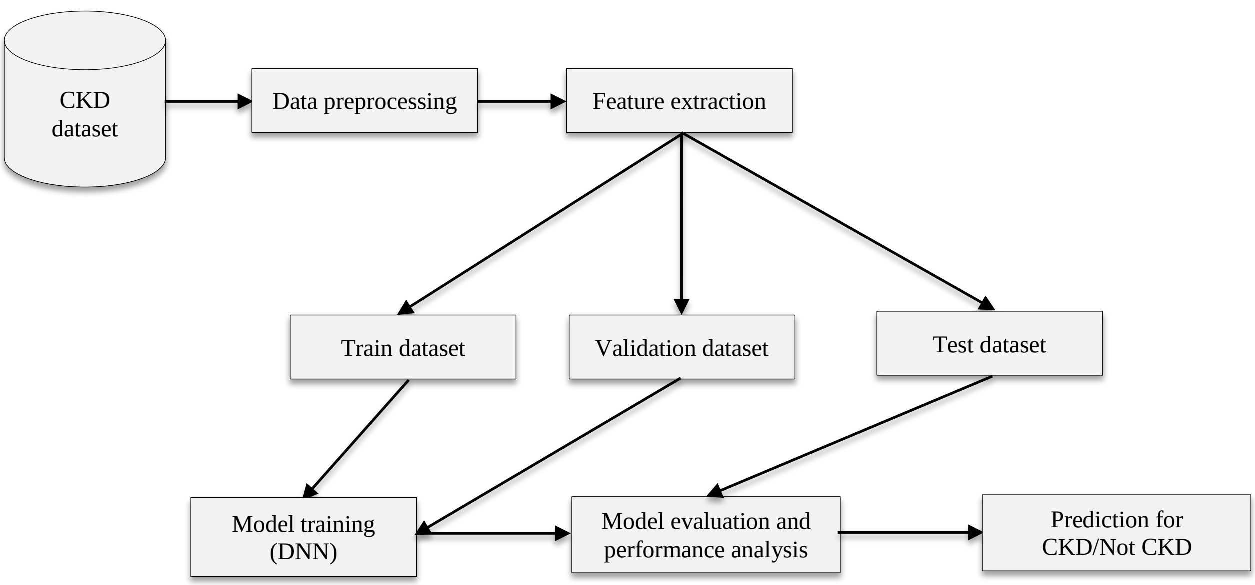


Figure 1 Complete of the proposed model

3.2 Data pre-processing

Clinical datasets in particular contain a lot of missing, noisy, redundant, and inconsistent data in today's real-world datasets. Results produced with low-quality data will also be low-quality. Every machine learning application starts with analyzing and interpreting the dataset to get it ready for modeling. This is known data preparation. Outliers are values that depart from the feature's overall trend. Incorrect outliers are caused by data input mistakes, which are referred to as data noise. When dealing with outliers, medical data cannot be treated like other data since the outliers might be actual (valid) or significant. As a result, each outlier discovered in CKD dataset is investigated to establish its plausibility. In this work use the mix min score to process the outliers. The analysis's extreme data points that fall outside of the range that is medically acceptable have been classified as missing data and then adjusted as shown in the section of missing data.

Missing values: Missing values are a common problem in real-world datasets, particularly in the medical industry. Typically, there are some missing values for every characteristic and patient record. In contrast, 96% of the variables in the chronic renal disease dataset have missing values; there is at least one missing value in 60.75% (243) of the cases, and 10% of all values are missing. Full null values in this research.

A feature selection in this investigation done by using the recursive feature elimination (RFE). RFE was introduced to select features of the subset that are not significant. Haemoglobin, SG, SC, red blood cell count, albumin, packed cell volume, and hypertension were found as key features in the RFE.

3.3 Proposed method

Dl model have been proposed to categorize CKD patients. The models' performance was evaluated using both all attributes and chosen features. The first step when used one dimensional (1D) to predicate dataset of CKD to classes CKD or not CKD is creating list of columns to retain and then transform non numeric columns into numerical columns also scaling feature by using min max scaler. The dataset is split into (80%) training, (10%) for validation and 10% for testing. This ratio gives the best result from other splitting when. Many experiments have been implemented to get the best result with changing some parameters such as the number of epochs, learning rate, and changing the data division ratio. In this section, will discuss the best-proposed model's design. A DNN model use in this work content of two convolutional layers of 1D, flattened layer, four dense layers. In this model, a dropout layer with a drop rate of 0.1 is placed after each dense layer.

This work tested different activation function on the CKD dataset and select the one that gave the best result which is rectified linear unit (ReLU) that replaces with sigmoid in the output layer.

Figure 2 presents the layered architecture of the proposed model. The other parameters are used in this model are Adam optimizer because obtained better result from other optimizers and binary cross entropy as loss because have a binary classification. The model training on 400 epochs with 8 batch sizes.

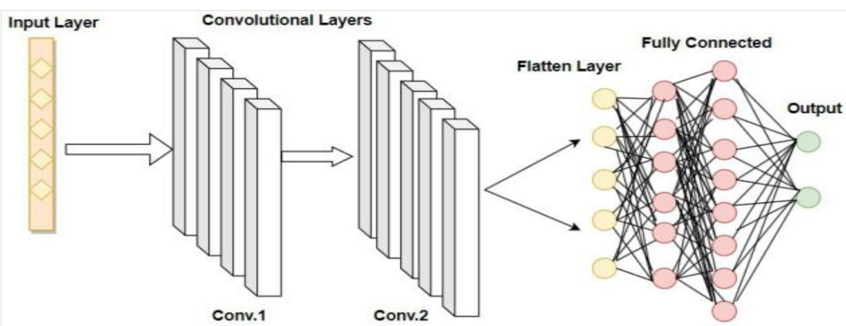


Figure 2 The architecture of the proposed model

4.Result

4.1Experiment setup

The studies are carried out using the Jupiter notebook application with anaconda and Python 3.9 is a computer programming language. Socket-learn, a free Python machine learning library, has been used to create a number of libraries. The processor was Intel cori7Gen9. Random access memory was 16GB.

4.2Evaluation parameters

Accuracy and F1-measure are the evaluation metrics considered in this work. While results of DL, the testing accuracy of the proposed model, testing loss, recall, precision, and F1-Score are all metrics used to evaluate this work achieving 99%, 98%, 98% for precision, recall, and F1-Score respectively. *Figure 3* shows the accuracy plot which achieved an accuracy of 99% for training and 98.33% for validation, as well as the loss decreased to 0.03 for training and less than 0.1 for validation. The time taken for model training is 235second. A binary classifier's confusion matrix was utilized to evaluate the performance by using expressions true positive (TP): are met when the expected value and the actual value are both positive, true negative (TN): are present when the data point's expected value and actual value are both negative, false positive (FP): These are cases where a data point's projected value is positive but its actual value was negative, and false negative (FN): are cases where a data point's expected value is negative while its actual value is positive[46]. *Figure 4* show the confusion matrix of the proposed model for two classes the (0) meaning CKD and the (1) meaning not CKD. The percentage of data that is in fact CKD and classified as CKD 44.33% (TP), while the error rate was 1.725(TN). The percentage of data that was in fact not CKD and was classified not CKD was 53.45% (FN), while the FP was 0%. The area of the AUC is defined by the bottom of the square and the receiver operating characteristic (ROC) curve. Areas under curve AUC values near 1 indicate strong performance, whereas values near 0.50 imply poor performance. *Figure 5* shows ROC/AUC curve of the proposed model that achieved 0.98.

4.3Comparative analysis of result

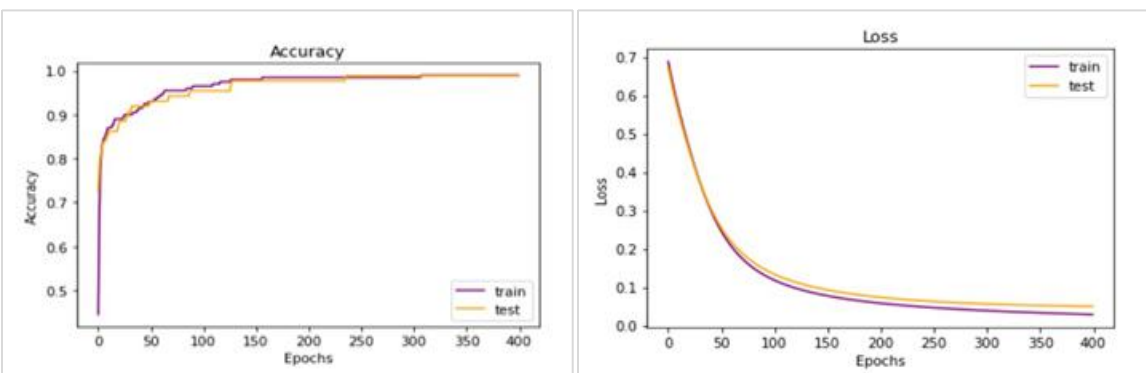
This section presents the results of the suggested model. There are three separate sets of CKD data sets: 80% for training, 10% for validation, and 10% for testing. *Table 2* displays the hyperparameter settings for the suggested model. The experimental outcomes of testing the suggested model using CKD data sets were displayed in *Tables 3* that illustrates the value of metrics for each class.

Table 2 Hyperparameter setting

Hyperparameter	Setting
Epochs	400
Bach size	8
Learning rate	0.1
Optimizer	Adam
Loss	Binary-cross entropy
Activation function	ReLU
Activation output layer	Sigmoid

Table 3 Metrics values and model classification report

	Precision	Recall	F1-Score	Support
0	0.97	100	0.99	39
1	1.00	0.95	0.98	23
Accuracy			0.98	62
Macro average	0.99	0.98	0.98	62
Weighted average	0.98	0.98	0.98	62



(a) (b)

Figure 3 Plot of the train and test accuracy with train and test loss

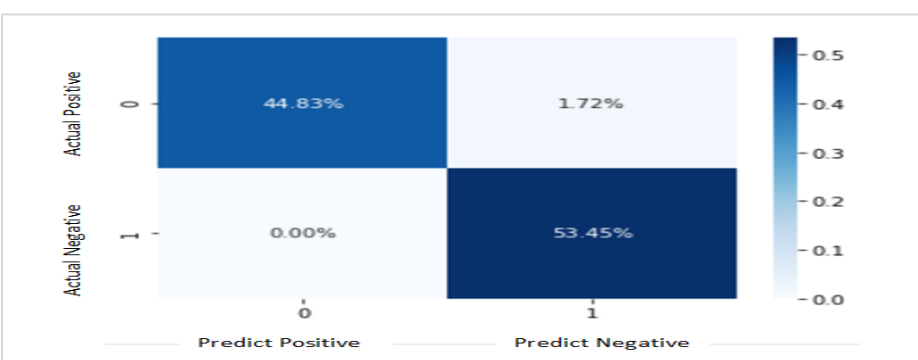


Figure 4 The proposed model's confusion matrix

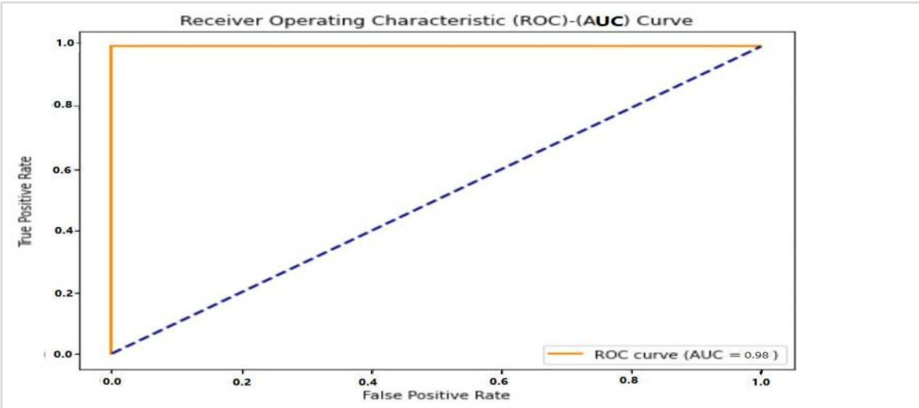


Figure 5 ROC-AUC curve of the proposed model

5.Discussion

The disease causes significant suffering and mortality, primarily due to early detection challenges and lack of awareness. Thus, early detection of CKD is considered crucial in halting the progression of the disease. Machine learning plays a vital role in early disease detection and prediction, aiding healthcare providers in making quick and accurate diagnoses.

In this study, DL methods were employed to predict the presence of chronic renal disease. The dataset comprised 24 features, including nominal and numerical values, as well as the class to which each instance belongs. The dataset had missing values, which were addressed during the data preparation stage.

Feature selection techniques and DL model were applied. The model evaluation process utilized metrics such as recall, F1-Score, accuracy, and precision. The confusion matrix displayed the correctly and incorrectly identified classes to assess the model's performance. The RFE technique was used for feature selection. The model's outcomes were compared with relevant research on the same dataset and classified. These findings are displayed in *Table 4.*

A limitation of the proposed model is that it has only been tested on limited datasets. In the future, significant volumes of more comprehensive and representative CKD data will be collected to diagnose disease severity and enhance model performance. The performance of the suggested model will also be evaluated using a large clinical dataset focusing on factors like acid-base parameters, hyperparathyroidism, inorganic phosphorus concentration, and nocturia. Additionally, new features will be incorporated to gain a broader understanding of the informative parameters associated with CKD, to assess prediction accuracy.

A complete list of abbreviations is summarised in *Appendix I.*

Table 4 Comparison this work with other works

Authors	Year	Dataset	Methodology	Result(Accuracy)
Salekin and Stankovic [31]	2016	The CKD dataset	K-NN, RF and ANN	98%
Debal and Sitote [32]	2022	The CKD dataset	DT, SVM	96.755
Xiao et al. [33]	2019	551 patients with CKD from the Department of Nephrology, Huadong Hospital	The following predictive models were developed and compared: LR, EN, lasso regression, RR, SVM, RF, XGBoost, CNN, and K-NN.	The model achieved an AUC and accuracy of 0.87 and 0.8, respectively, and incorporated various techniques such as elastic net, lasso regression, RR, and LR. Among these, LR performed the best with AUCs of 0.873, 0.83, and 0.82. Elastic Net demonstrated the highest sensitivity at 0.85 and the best specificity at 0.83.
Islam et al. [34]	2023	The CKD dataset	10-fold cross validation	98%
This work	2023	The CKD dataset	DNN	98.33%

6.Conclusion and future work

The aim of this research is to demonstrate the effectiveness of DL in diagnosing CKD using the fewest possible tests or features. An analysis of the relationships between variables was conducted to minimize the number of features and eliminate redundancy. After applying a filter feature selection method to the remaining characteristics, hemoglobin, albumin, and SG were identified as having the greatest impact on predicting CKD development. The purpose is to validate the findings with a larger dataset or to compare them with a dataset of similar quality in future research, considering that this study utilized a limited amount of data. Additionally, with relevant data, it may be possible to predict an individual's likelihood of developing CKD in the future if they have conditions such as diabetes, high blood pressure, or a family history of renal failure. The DNN model achieved an accuracy of 98.33% and an F1-Score of 98.41%. DL algorithms are considered reliable techniques that can aid in accelerating the diagnosis process of various diseases in medical imaging. The unpredictability of CKD, due to its lack of dependence on a single trait, poses a challenge in diagnosis. Moreover, common symptoms of CKD do not significantly contribute to its diagnosis. Future research projects can employ clustering approaches to enhance classification accuracy and reduce instances of misclassification.

Acknowledgment

None.

Conflicts of interest

The authors have no conflicts of interest to declare.

Data Availability

The dataset used in this study is publicly available at https://archive.ics.uci.edu/ml/datasets/Chronic_Kidney_Disease

Author’s contribution statement

Ashwan A. Abdulmunem: Conceptualization and design of the work, implementation, and analysis. **Alaa Jamal Jabbar:** Conceptualization and design of the work, implementation, and analysis.

References

[1] Altalbe A, Javed AR. Applying customized convolutional neural network to kidney image volumes for kidney disease detection. Computer Systems Science & Engineering. 2023; 47(2):2119-34.

[2] Houssein EH, Sayed A. A modified weighted mean of vectors optimizer for chronic kidney disease classification. Computers in Biology and Medicine. 2023; 155:106691.

[3] Sisodia A, Jindal R. An effective model for healthcare to process chronic kidney disease using big data processing. Journal of Ambient Intelligence and Humanized Computing. 2023; 14(10):1-7.

[4] Ye Z, An S, Gao Y, Xie E, Zhao X, Guo Z, et al. The prediction of in-hospital mortality in chronic kidney disease patients with coronary artery disease using machine learning models. European Journal of Medical Research. 2023; 28(1):1-13.

[5] Farjana A, Liza FT, Pandit PP, Das MC, Hasan M, Tabassum F, et al. Predicting chronic kidney disease using machine learning algorithms. In 13th annual computing and communication workshop and conference 2023 (pp. 1267-71). IEEE.

[6] Harimoorthy K, Thangavelu M. Multi-disease prediction model using improved SVM-radial bias technique in healthcare monitoring system. Journal of Ambient Intelligence and Humanized Computing. 2021; 12:3715-23.

[7] Moreno-sánchez PA. Data-driven Early diagnosis of chronic kidney disease: development and evaluation of an explainable AI model. IEEE Access. 2023; 11:38359-69.

[8] Amirgaliyev Y, Shamiluulu S, Serek A. Analysis of chronic kidney disease dataset by applying machine learning methods. In 12th international conference on application of information and communication technologies 2018 (pp. 1-4). IEEE.

[9] Bello AK, Levin A, Lunney M, Osman MA, Ye F, Ashuntantang GE, et al. Status of care for end stage kidney disease in countries and regions worldwide: international cross sectional survey. BMJ. 2019; 1-13.

[10] Mitani A, Hammel N, Liu Y. Retinal detection of kidney disease and diabetes. Nature Biomedical Engineering. 2021; 5(6):487-9.

[11] Vijayalakshmi A, Sumalatha V. Survey on diagnosis of chronic kidney disease Using Machine learning algorithms. In 3rd international conference on intelligent sustainable systems 2020 (pp. 590-5). IEEE.

[12] Twarish Alhamazani K, Alshudukhi J, Aljaloud S, Abebaw S. Implementation of machine learning models for the prevention of kidney diseases (CKD) or their derivatives. Computational Intelligence and Neuroscience. 2021; 2021:1-9.

[13] Fabian J, George JA, Etheredge HR, Van DM, Kalyesubula R, Wade AN, et al. Methods and reporting of kidney function: a systematic review of studies from sub-Saharan Africa. Clinical Kidney Journal. 2019; 12(6):778-87.

[14] Chittora P, Chaurasia S, Chakrabarti P, Kumawat G, Chakrabarti T, Leonowicz Z, et al. Prediction of chronic kidney disease-a machine learning perspective. IEEE Access. 2021; 9:17312-34.

[15] Padmanaban KA, Parthiban G. Applying machine learning techniques for predicting the risk of chronic kidney disease. Indian Journal of Science and Technology. 2016; 9(29):1-6.

[16] Chimwayi KB, Haris N, Caytiles RD, Iyengar NC. Risk level prediction of chronic kidney disease using

neuro-fuzzy and hierarchical clustering algorithm(s). International Journal of Multimedia and Ubiquitous Engineering. 2017; 12(8):23-6.
[17] Sandeep J, Mula R, Kuriachan B. Chronic kidney disease analysis using machine learning algorithms. International Journal for Research in Applied Science & Engineering Technology. 2018; 6(1):3367-79.
[18] Lambert JR, Perumal E. Oppositional firefly optimization based optimal feature selection in chronic kidney disease classification using deep neural network. Journal of Ambient Intelligence and Humanized Computing. 2022; 13(4):1799-810.
[19] Almasoud M, Ward TE. Detection of chronic kidney disease using machine learning algorithms with least number of predictors. International Journal of Soft Computing and Its Applications. 2019; 10(8): 89-96.
[20] Lee HC, Yoon HK, Nam K, Cho YJ, Kim TK, Kim WH, et al. Derivation and validation of machine learning approaches to predict acute kidney injury after cardiac surgery. Journal of Clinical Medicine. 2018; 7(10):1-13.
[21] Kosarkar N, Basuri P, Karamore P, Gawali P, Badole P, Jumle P. Disease prediction using machine learning. In 10th international conference on emerging trends in engineering and technology-signal and information processing 2022 (pp. 1-4). IEEE.
[22] Sidey-gibbons JA, Sidey-gibbons CJ. Machine learning in medicine: a practical introduction. BMC Medical Research Methodology. 2019; 19:1-8.
[23] Song W, Liu Y, Qiu L, Qing J, Li A, Zhao Y, et al. Machine learning-based warning model for chronic kidney disease in individuals over 40 years old in underprivileged areas, Shanxi Province. Frontiers in Medicine. 2023; 9:930541.
[24] Song S, Yuan B, Zhang L, Cheng G, Zhu W, Hou Z, et al. Increased inequalities in health resource and access to health care in rural China. International Journal of Environmental Research and Public Health. 2019; 16(1):1-10.
[25] Wang K, Tian J, Zheng C, Yang H, Ren J, Li C, et al. Improving risk identification of adverse outcomes in chronic heart failure using SMOTE+ ENN and machine learning. Risk Management and Healthcare Policy. 2021: 2453-63.
[26] Zhang K, Liu X, Xu J, Yuan J, Cai W, Chen T, et al. Deep-learning models for the detection and incidence prediction of chronic kidney disease and type 2 diabetes from retinal fundus images. Nature Biomedical Engineering. 2021; 5(6):533-45.
[27] Ma F, Sun T, Liu L, Jing H. Detection and diagnosis of chronic kidney disease using deep learning-based heterogeneous modified artificial neural network. Future Generation Computer Systems. 2020; 111:17-26.
[28] Tapia-conyer R, Gallardo-rincón H, Betancourt-cravioto M. Chronic kidney disease in disadvantaged populations: online educational programs for NCD prevention and treatment. In chronic kidney disease in disadvantaged populations 2017 (pp. 337-45). Academic Press.
[29] Krishnamurthy S, Ks K, Dovgan E, Luštrek M, Gradišek PB, Srinivasan K, et al. Machine learning prediction models for chronic kidney disease using national health insurance claim data in Taiwan. Healthcare 2021; 9(5):1-13.
[30] Sabanayagam C, Xu D, Ting DS, Nusinovici S, Banu R, Hamzah H, et al. A deep learning algorithm to detect chronic kidney disease from retinal photographs in community-based populations. The Lancet Digital Health. 2020; 2(6):e295-302.
[31] Salekin A, Stankovic J. Detection of chronic kidney disease and selecting important predictive attributes. In IEEE international conference on healthcare informatics (ICHI) 2016 (pp. 262-70). IEEE.
[32] Debal DA, Sitote TM. Chronic kidney disease prediction using machine learning techniques. Journal of Big Data. 2022; 9(1):1-9.
[33] Xiao J, Ding R, Xu X, Guan H, Feng X, Sun T, et al. Comparison and development of machine learning tools in the prediction of chronic kidney disease progression. Journal of Translational Medicine. 2019; 17(1):1-3.
[34] Islam MA, Majumder MZ, Hussein MA. Chronic kidney disease prediction based on machine learning algorithms. Journal of Pathology Informatics. 2023; 14:100189.
[35] Vinay R, Soujanya KL, Singh P. Disease prediction by using deep learning based on patient treatment history. International Journal of Recent Technology and Engineering. 2019; 7(6):745-54.
[36] Qadir AM, Abd DF. Kidney diseases classification using hybrid transfer-learning densenet201-based and random forest classifier. Kurdistan Journal of Applied Research. 2023:131-44.
[37] Venkatrao K, Kareemulla S. HDLNET: a hybrid deep learning network model with intelligent IOT for detection and classification of chronic kidney disease. IEEE Access. 2023; 11:99638-52.
[38] Alikhan JS, Alageswaran R, Amali SM. Self-attention convolutional neural network optimized with season optimization algorithm espoused chronic kidney diseases diagnosis in big data system. Biomedical Signal Processing and Control. 2023; 85:105011.
[39] Singh V, Jain D. A hybrid parallel classification model for the diagnosis of chronic kidney disease. International Journal of Interactive Multimedia and Artificial Intelligence. 2023; 8(2):14-28.
[40] Liang P, Yang J, Wang W, Yuan G, Han M, Zhang Q, et al. Deep learning identifies intelligible predictors of poor prognosis in chronic kidney disease. IEEE Journal of Biomedical and Health Informatics. 2023; 27(7): 3677-85.
[41] Kaur C, Kumar MS, Anjum A, Binda MB, Mallu MR, Al AMS. Chronic kidney disease prediction using machine learning. Journal of Advances in Information Technology. 2023; 14(2):1-8.
[42] Busi RA, Stephen MJ. Effective classification of chronic kidney disease using extreme gradient boosting algorithm. International Journal of Software Innovation. 2023; 11(1):1-8.

[43] Lu Y, Ning Y, Li Y, Zhu B, Zhang J, Yang Y, et al. Risk factor mining and prediction of urine protein progression in chronic kidney disease: a machine learning-based study. BMC Medical Informatics and Decision Making. 2023; 23(1):173.

[44] Almansour NA, Syed HF, Khayat NR, Altheeb RK, Juri RE, Alhiyafi J, et al. Neural network and support vector machine for the prediction of chronic kidney disease: a comparative study. Computers in Biology and Medicine. 2019; 109:101-11.

[45] Maisha SJ, Biswangri E, Hossain MS, Andersson K. An approach to detect chronic kidney disease (CKD) by removing noisy and inconsistent values of UCI dataset. In proceedings of the third international conference on trends in computational and cognitive engineering: TCCE 2021 2022 (pp. 457-72). Singapore: Springer Nature Singapore.

[46] Bhattacharya M, Jurkovitz C, Shatkay H. Chronic kidney disease stratification using office visit records: handling data imbalance via hierarchical meta-classification. BMC Medical Informatics and Decision Making. 2018; 18:35-44.

Ashwan A. Abdulmunem received her Ph.D. degree in Computer Vision and Artificial Intelligence from Cardiff University, UK. She is currently an Assistant Professor at the College of Computer Science and Information Technology, University of Kerbala. Her research interests encompass Computer Vision and Graphics, Computational Imaging, Pattern Recognition, Artificial Intelligence, as well as Machine and Deep Learning.
Email: ashwan.a@uokerbala.edu.iq

Alaa Jamal Jabbar received her B.Sc. degree in Computer Science from Thiqar University, Iraq. She has a Master degree in computer science at the General Directorate of Education of Dhi Qar. Her research interests include Computer Vision, Artificial Intelligence, as well as Machine and Deep Learning.
Email: rdaalzydy00@gmail.com

Appendix 1

S. No.	Abbreviation	Description
1	1D	One Dimensional
2	2D	Two Dimensional
3	ANN	Artificial Neural Network
4	AO	Aquila Optimization
5	AUC	Area Under the Curve
6	BG	Blood Glucose
7	BP	Blood Pressure
8	BU	Blood Urea
9	Caps Net	Capsule Network
10	CKD	Chronic Kidney Disease
11	CT	Computational Thinking
12	DLA	Deep Learning Application
13	DNN	Deep Neural Network
14	DT	Decision Tree
15	DL	Deep Learning
16	DSCNN	Deep Separable Convolution Neural Network
17	ESRD	End Stage Renal Disease
18	FP	False Positive
19	FN	False Negative
20	GFR	Glomerular Filtration Rate
21	ISSA	Improved Salp Swarm Algorithm
22	IoT	Internet of Things
23	K-NN	k-Nearest Neighbors
24	LR	Logistic Regression
25	LASSO	Least Absolute Shrinkage and Selection Operator
26	NB	Naive Bayes
27	PCA	Principal Component Analysis
28	PC	Pus Cell
29	PCC	Pus Cell Clumps
30	RBC	Red Blood Cells
31	ReLU	Rectified Linear Unit
32	RF	Random Forest
33	RFE	Recursive Feature Elimination
34	ROC	Receiver Operating Characteristic
35	RMSE	Root Mean Square Error
36	RR	Ridge Regression
37	SACNN	Self-Attention Convolutional Neural Network
38	SC	Serum Creatinine
39	SG	Specific Gravity
40	SMOTE	Synthetic Minority OverSampling Technique
41	SOA	Season Optimization Algorithm
42	STOA	Sooty Tern Optimization Algorithm
43	SVC	Support Vector Cluster
44	SVM	Support Vector Machine
45	TP	True Positive
46	TN	True Negative
47	UCI	University of California Irvine