Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2023) 31: 951 – 968

doi:10.55730/1300-0632.4028

Hybrid machine learning model to predict chronic kidney diseases using handcrafted features for early health rehabilitation

Amjad Rehman[1], **Tanzila Saba**[1*], **Haider Ali**[2], **Narmine El Hakim**[1], **Noor Ayesha**[3]
[1]AIDA Lab CCIS, Prince Sultan University, Riyadh, Saudi Arabia
[2]Department of Statistics, University of Gujrat, Gujrat, Pakistan
[3]School of Clinical Medicine, Zhengzhou University, Zhengzhou, Henan, RP China

Received: 22.04.2023 • **Accepted/Published Online:** 16.08.2023 • **Final Version:** 27.10.2023

Abstract: Chronic kidney diseases proliferate due to hypertension, diabetes, anemia, obesity, smoking etc. Patients with such conditions are sometimes unaware of first symptoms, complicating disease diagnosis. This paper presents chronic kidney disease (CKD) prediction model to classify CKD patients from NCKD (Non-CKD). The proposed study has two main stages. First, we found the odds ratio through logistic regression and comparison test to identify early risk factors from kidneys' MRI and differentiate CKD from NCKD subjects. In stage 2, LR, LDA, MLP classifiers were applied to predict CKD and NCKD by extracting features from MRI. The odds ratio of blood glucose random and serum creatinine was found higher, and levels of sodium, Potassium, packed cell volume, white blood cell count, and red blood cell count were found lesser in CKD. The comparison results show increase levels in blood glucose random, serum creatinine and decreased levels found in sodium, potassium, packed cell volume, White blood cell and red blood cell count respectively in CKD patients than NCKD subjects. The accuracies of LR were 98.5% and 97.5% for train & test datasets. While LDA accuracy was 96.07% and 96.6% for train and test datasets. Likewise, MLP attained were 95% and 94.1% accuracy for train and test datasets. Finally, we used 5-fold CV approach on the LR model. The mean accuracies of LR were 0.954 and 0.942 for training and testing data respectively. According to LR the serum creatinine, Albumin, Diabetes mellitus, red blood cells count, pus cell and hypertension were found to be the most significant features to discriminate the CKD patients from NCKD. The proposed strategy is best suited for practical implementation for reducing the disease's prevalence.

Key words: Health rehabilitation, chronic kidney disease (CKD), prediction, public health, health risks

1. Introduction

Chronic Kidney Disease (CKD) is a long-term condition where the kidneys lack the essential effort to function. Around 1.2 million individuals died in 2018, yet only 5-12 percent were due to CKD [1, 2]. Patients and their families confront several financial and ethical concerns due to the high prevalence of CKD, high treatment costs, and variable treatment accessibility. Creatinine is a chemical waste product produced by muscle metabolism to diagnose CKD. The typical range for women is 1.2mg/dl, and for males is 1.46mg/dl, however, later stages of CKD generate more creatinine [3, 4]. Chronic noncommunicable diseases (NCD) do not spread from person to person, but communicable diseases (CD) do. The consequences of chronic illness may be fatal. This condition affects people of all ages worldwide. The kidney is a critical organ in the human body, and its diseases are

*Correspondence: tsaba@psu.edu.sa

chronic. A kidney's primary purpose is to filter the blood utilising millions of nephrons to eliminate undesirable substances from the body—chronic renal disease results from not excreting undesirable materials [5]. The National Kidney Foundation (NKF) [6] recently reported that CKD mortality is more significant than breast or prostate cancer. Only in the USA 37 million individuals are expected to be unaware about CKD. Therefore, around 80 million individuals are at risk for CKD. COVID-19 has recently been linked to significant renal disease and transplant difficulties [7].

Two kidneys, which are essential organs for the body's healthy operation, are placed in the peritoneal cavity in the rear of the human body. The primary job of the kidneys is to maintain a proper level of water, salt, and other ions and trace elements in the body, including acids, calcium, phosphorus, magnesium, potassium, chlorine, and trace amounts of other elements. The kidneys also release hormones including erythropoietin, vitamin D, and renin at the same time. Erythropoietin primarily promotes the development and maturation of red blood cells in the bone marrow, whereas vitamin D controls the body's levels of calcium and phosphorus, as well as bone structure and many other processes. Additionally, hormones that control blood pressure, fluid balance, bone metabolism, and vascular calcifications work through the kidneys. Finally, all of the metabolic waste products, medications, and other poisons that enter the body are removed by the kidneys [8]. The two primary causes of chronic kidney disease are diabetes and high blood pressure. High blood sugar levels are a hallmark of diabetes, which affects the kidneys, heart, blood vessels, and eyes. Furthermore, the inability to regulate high blood pressure can significantly increase the risk of heart attacks, strokes, and chronic renal disease. Glomerulonephritis, genetic illnesses, dysplasia, kidney stones, tumours, recurrent urinary tract infections, metabolic diseases, obesity, and ageing are other kidney-related problems [9].

Expert intelligent systems may be used to identify CKD warning symptoms. Data mining technologies may classify data based on many criteria to detect diseases and retrieve other pertinent data. Expert systems can make an automated prediction based on patient data to deliver improved clinical outcomes to practitioners [10]. Due to multi-dimensional data processing, expert systems have outperformed human professionals in several circumstances. To extract information from a database and uncover comparable data for decision-making, evaluation, and forecasting, data mining methods are applied [11]. Data mining uses descriptive and predictive models. Describing patterns in data, explanatory models classify patterns in data, whereas predictive models predict outcomes from multiple data sources. There are two types of data mining models: predictive and time-sequence models. They use classification and clustering to corroborate the accurate forecast. The data state might exacerbate the challenge, for example, noise, missing labels, and dynamic and big data sets. Data issues reduce the performance of machine learning algorithms [12]. When dealing with medical data sets, similar challenges arise. Common bioinformatics difficulties need to be addressed in ordinary diagnostics. Practitioners suggest several tests to learn more about the ailment for proper diagnosis and the lack of comprehensive testing frequently complicates the diagnosing procedure. Additionally, several tests may delay proper diagnosis, increase treatment costs, and impair prediction accuracy. Machine learning algorithms may easily overcome such shortcomings, since they are trained with small datasets in comparison to the deep learning models [13, 14]. Furthermore, machine learning algorithms are widely employed in healthcare to diagnose and predict infectious diseases at early stage. Another feature that makes machine learning techniques popular among practitioners is their capacity to handle complicated and extensive datasets [15, 16]. Healthcare academics and industry use cutting-edge statistical tools to help and even direct practitioners. Statistical approaches may manually create or automatically extract features from medical data. Quality of outcomes relies on the quality of employed features, which in turn depends on the proposed algorithm's ability [17].

2. Background

Growing medical facilities and a growing human population have contributed to more biological data. Confoundingly, disease-related mortality is growing. Early health detection failure is one of the key reasons for higher mortality. However, certain diseases are significantly harder to diagnose early [18]. Like kidney failure, cancer, heart disease, asthma, and diabetes, CKD is a slow-progressing condition. Diverse classification challenges to predict chronic diseases have been completed recently. Classification algorithms can also forecast CKD patients' conditions, increasing prediction accuracy. Many studies have been published recently to improve clinical diagnosis accuracy. An early diagnosis would allow for a better cure. Support vector machine (SVM), Artificial neural network (ANN), K-nearest neighbor (KNN), and random forest (RF) are popular machine learning algorithms in AI-based health care. Diabetes leads to heart disease, kidney failure, blindness and strokes [19, 20].

An algorithm for predicting diabetes patients' risk of heart disease has been developed in. They used clinical data from the online repository and the campus medical institution to conduct experiments. The method predicts eight diagnostic attributes for diabetes. Naive Bayes beat J48 by 79.56 percent on their dataset. Dovgan et al., [16] proposed ML models to forecast the need for renal replacement therapy for CKD patients. They got a 0.773 area under the curve (AUC) for predicting renal replacement therapy within 12 months after CKD diagnosis using data from 8,492 individuals. They further concluded that machine learning algorithms developed could be used as a feasible screening tool for predicting the period in which a patient with chronic kidney disease may need renal replacement treatment.

Polat et al.[21] presented an SVM-based technique to enhance medical systems' performance and minimise fatality rates. They obtained 400 instances with 24 attributes from the UCI machine repository using free source datasets available. Feature selection strategies are used to decrease data dimensionality. The classifier predicted the presence or absence of CKD using binary outcomes. It was 98.5 percent accurate using 10-fold cross validation and SVM with the best initial search. Ahmad et al. [22] discussed several kidney diseases, symptoms, and risk factors. They employed Naive Bayes, KNN, and Logistic Regression (LR) to forecast kidney diseases that improve performance. Classification modelling and expert system development were applied. The system's development includes data gathering, preprocessing, and classification steps. The suggested system has a reported accuracy of 98.34. A study by Rodrigues et al. looked at the effects of CAP dialysis on kidney patients. They created a database comprising 850 patients' records. This dataset was created over 8 years. Kidney dialysis treatment uses Naive Bayes, KNN, LR, MLP, and RT classifier algorithms. According to this research, K-NN attained 99.65 percent accuracy.

In [23], authors used computational approaches to detect depression from speech analysis. There are 18 hand created features in this dataset. This research aims to predict patient survival. They employed Naive Bayes, ANNs, and Decision Trees as algorithms. These approaches outperformed other methods in terms of accuracy and time complexity. With 84% accuracy in prediction, Naive Bayesian takes 0.01 seconds. They found that selecting features from the Hepatitis patient data set enhanced prediction accuracy. Multilayer Perceptron, Sequential Minimum Optimization, J48 and Naive Bayesian algorithms were employed in the BFS and Greedy Search based CFS. It is based on BFS and GS. Naive Bayesian feature choices in BFS and GS boost accuracy by 80% in 0.01 seconds. Using classification algorithms, study [24] predicted kidney disease progression phases. Normal kidney function, Acute Nephritic Syndrome, Chronic Kidney Disease, Acute Renal Failure, and Chronic Glomerulonephritis. Naive Bayes (NB) and Support Vector Machines (SVM) are employed (SVM). To predict kidney disease, six variables were used: class, gender, age, urea, serum creatinine, and

Glomerular Filtration Rate. The classification employed a dataset of 584 cases with six attributes from various data collecting locations such as medical laboratories and hospitals. In these datasets, two nominal and four numerical attributes diagnose the illness. Machine learning algorithms can perform adequately on this sized dataset even if little. SVM and Naive Bayes achieved an accuracy of 76.32% and 70.96%. SVM finished in 3.22 seconds and Naive Bayes in 1.29 seconds, respectively. For these datasets, SVM outperformed Naive Bayes. Abraham et al. [19] tested machine learning models for kidney stone formation from urine data. They trained XGBoost and logistic regression to forecast stone content using 24H urine samples. Projections show either binary (calcium versus noncalcium stone) or multiclass (calcium oxalate, uric acid, hydroxyapatite, or other) stone types. They discovered task-specific predictors and assessed performance using ROC-AUC and accuracy. With a ROC-AUC of 0.80 for both models, XG beat LR in terms of identifying binary stone structure (91% vs 71%).

Risk assessment in clinical practise for patients with CKD was shown to be reliable by [25, 26]. This study utilised the UC Irvine Chronic Renal Failure (CFR) database. Three different types of independent modelling of class analogy—KNN, SVM, and Soft—were utilised in their analysis. The SVM model eliminated more background noise than either of the other two models. The SVM achieved a 99% accuracy in this test. The author of [27] created a technique to help physicians predict the incidence of CRF in their patients. Using the CRF dataset housed in the UCI repository, the authors classified the data using KNN, Naive Bayes, LDA, random subspace, and tree-based methods. Researchers found that a KNN classifier applied to a random subspace had a 94% success rate. Similar decision assistance to [28] was developed by the authors of another research [28]. In this research, the authors employ ANN, Naive Bayes, and decision tree methods to categorise CRF. Different machine learning algorithms were tested, and their results were compared, using data collected from Prince Hamza Hospital in Amman, Jordan. When compared to two other methods, the decision tree is anticipated to yield the most accurate results. Using electronic health record (EHR) and billing data from diabetic patients, Schober et al. [29] developed a prediction model based on gradient boosting to identify CKD. The authors of [30] employed SVM, decision trees, Nave Bayes, and KNN to identify CKD in data sets from the University of California, Irvine. In order to choose characteristics, the authors created a ranking system. The decision tree scored 99.75, more than any of the other three machine learning approaches. In [31], the authors of a hierarchical multiclass classification method for chronic renal illness detection in an imbalanced data set were described.

According to Azar et al. [32], kidney failure progresses with time. It detects kidney failure progression and saves time using the Adaptive Neuro-Fuzzy Inference System (ANFIS). The hospital's 10-attribute dataset is utilised to predict kidney disease. An example of a uric acid test is a urine test. Derived from 465 cases, 277 were male. ANFIS reports a 95% accuracy rate. They claim that machine learning may assist diagnose chronic kidney illness and aid in clinical practice. The author uses algorithms like SVM, KNN, and DT to diagnose CKD in patients. The dataset is split 70:30 across training and testing sets. From the tested algorithms, SVM has the most fantastic accuracy at 98.3 % (average accuracy utilising fivefold cross-validation). Table 1 demostrate the pros and cons of literature review.

Based on the literature explored, it is highlighted that current kidney diagnosis solutions inefficient in identifying kidney failure at their early stage. At the same time deep learning models are computational expensive as well as need large datasets to train while machine learning models have the ability to generalize from small datasets. Hence, this research has the following main contributions.

- Developed and validated a machine learning model to predict chronic kidney disease patients from MRI for early health rehabilitation.
- Identified leading causes of kidney failure factors by evaluating MRI of kidneys for early health rehabilitation.
- The proposed model could be used for chorionic kidney disease diagnosis in remote areas to support physicians and patients.

Further research comprises three main sections, section 3 presents the proposed methodology, and section 4 exhibits results and discussion. Finally, section 5 summarizes the research.

3. Proposed Methodology

This study is conducted to identify the most risk factors of CKD by comparing risk factors between CKD & NCKD and develop the prediction models to classify CKD patients for early health rehabilitation. The logistic regression was used for the odds ratio and as well as for prediction. The t-test was used for the comparison of risk factors of both the CKD and NCKD group. The classification process is performed using machine learning models such as LR, LDA, and deep learning models such as MLP. The other Machine learning models did not achieve a higher accuracy rate than these models, so other machine learning models were excluded from the experiment. The other deep learning models that were not used, such as Convolution neural network, recurrent neural network, etc., have their properties and limitations regarding the data nature. The benchmark dataset employed in this research is detailed in the next section. The data consist of 25 variables, where group variables are used as the dependent variable, and all other variables were used as independent variables. We divide our work into two stages in first stage, we find the odds ratio through logistic regression and determine which effects are significant. The odds ratio provides the sensitivity analysis of interest between the dependent and independent variables. If the odds ratio <1 means, that is a decreased likelihood of an event occurrence. If >1 means that is an increased likelihood of an event occurrence. If the odds ratio=1 both groups have the same effect [28]. The odds and odds ratio can find from the logistic regression is an important concept related to the LR . The LR also is used for classification rather than the regression model only. It is very efficient, performs well, and is a simple binary dependent variable classification problem model. Then significant variables are further compared with the T-test to know how much risk factors are different in CKD than NCKD. The T-test is commonly used to compare the mean of two groups. It draws how the mean of a group is different from each other. While stage 2 predicted the CKD and NCKD subjects using machine learning classifiers such as LR, LDA, and Deep learning models such as MLP. All models were designed with their parameters and set the 94% above convergence value to select the best design or model for prediction. The best-selected design was used for testing samples to check the high accuracy, and this process was repeated until high accuracy was also not obtained for testing samples. Figure 1 presents the proposed architecture. The LDA is a popular technique used for pattern recognition and classification. It is used in different applications for classification purposes, such as the medical field, pneumonia diagnosis classification, lung cancer classification, electromyography (EMG) signals analysis, etc [33].

The MLP Artificial Neural network (ANN) proceeds the information based on the biological nervous system. The ANN helps in solving a large number of real-world problems. The ANNs have various architectures. However, in this study, we used MLP for features extraction from kidney MR images and forecast into CKD and NCKD. The MLP is considered an example of a feedforward neural network. It has three layers input,

Table 1. Pros and Cons of literature review

References	Methodology	Aim	Pros	Cons
[17]	svm	prediction of kidney stone composition using information acquired from electronic health records	Authors utilized many machine learning models for chronic disease prediction.	Extensive preprocessing and automatic deep learning architecture was not utilized.
[19]	Naïve bayes	Chronic disease prediction with machine learning	They used clinical data from the online repository and the campus medical institution to conduct experiments.	The proposed method was not accurate.
[14]	ML	Machine learning algorithms were used to predict the commencement of renal replacement therapy in individuals with chronic kidney disease.	They got a 0.773 area under the curve (AUC) for predicting renal replacement therapy within 12 months after CKD diagnosis using data from 8,492 individuals.	They achieved low results for CKD predictions and dataset utilized was limited cause overfitting.
[16]	SVM	Predict kidney disease progression phases	To predict kidney disease, six variables were used: class, gender, age, urea, serum creatinine, and Glomerular Filtration Rate. The classification employed a dataset of 584 cases with six attributes from various data collecting loca- tions such as medical laboratories and hospitals	Limited dataset and accuracy is low.
[18]	DT	Kidney failure progresses with time.	It detects kidney failure progression and saves time using the Adaptive Neuro-Fuzzy Inference System (ANFIS)	Authors was not employed any deep learning to enhance the accuracy.
[19]	SVM	Several feature selection techniques to diagnose chronic renal disease.	Various feature selection strategies are used to decrease data dimensionality. The classifier predicted the presence or absence of CKD using binary outcomes	Only employed ML models with k-fold cross validation.
[20]	LR	Forecast kidney diseases that improve performance.	They employed Naive Bayes, KNN, and Logistic Regression (LR) to forecast kidney diseases that improve performance.	The study did not propose anything new; state-of-the-art ML models were utilized.

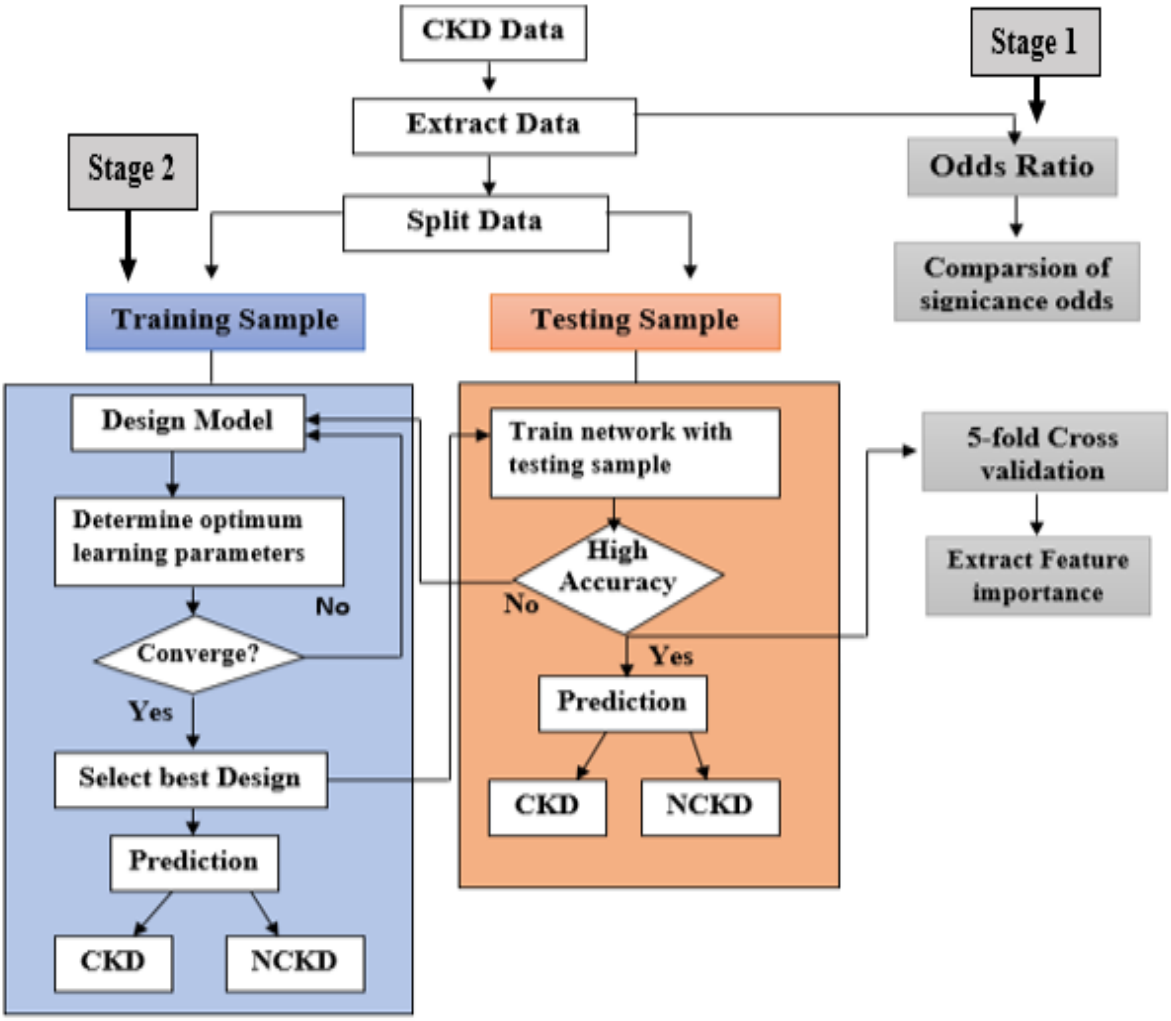

Figure 1. Proposed research framework

hidden layers, and output layers, and the one layer connects with to next layer. The MLP maps the input data on output categories appropriately [34]. The modified architecture of the MLP model is exhibited in Figure 2. The training /testing approach is used for prediction where 70% part of data is used as training and 30% is used as testing. The ML and DL models trained on the training part and tested on the test dataset to determine the prediction error and validity of the models. The performance of the classifiers was checked through the accuracy, sensitivity, specificity, ROC, Area under the curve (AUC), false-positive rate (FPR), and false-negative rate (FNR).

3.1. Data description

The dataset was used in this study downloaded from https://archive.ics.uci.edu/ml/index.php. The University of California, Irvine's online machine learning repository, gets the data from chronic illness disease from a health care unit within two months. The data consist of 25 variables, where 11 are in numeric form and 14 are in

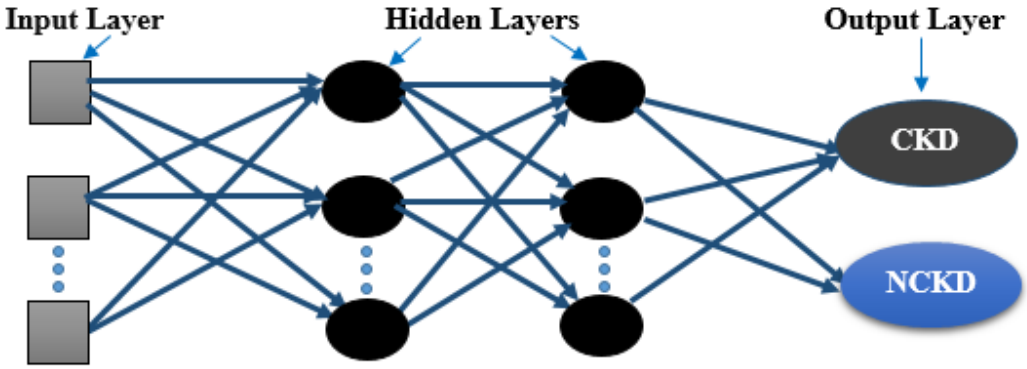

Figure 2. Modified MLP architecture.

Algorithm for Proposed Methodology

Input: CKD Dataset
Output: Prediction
Start:
1. Extract data features = comparison of significance odds ⟵ Odds ratio
2. Data splitting
   Train
   Test
3. Training samples
4. Design the model = Extract best features ⟵ 5fold cross validation
5. Determine the optimum learning parameters
6. Select best design
7. Train network with test samples
8. Prediction
   CKD
   NCKD

End

nominal form with binary and multiple labels. The demographic information was collected from the subjects, such as Age, weight, and gender. Obtain the communication of cells such as red blood cells, pus cells, pus cell clumps, number of single-cell bacteria, packed cells volume, and white blood cell count. Other information such as blood pressure, specific gravity, presence of coronary artery disease, and presence of diabetes, hypertension, diabetes mellitus, albumin, random blood glucose, urea, serum creatinine, sodium, potassium, hemoglobin, appetite, pedal edema, anemia, and variable contains the class of CKD and NCKD.

3.2. Tool Description

All the analysis were performed on Google Colab using the Graphics processing units (GPU). The Google Colab provide the facility to do python analysis fast through GPU for 12 hours. The proposed models were performed under GPU where LR, LDA, and MLP took few seconds but LR take less time than LDA and MLP model.

The proposed model algorithm is presented below

Input: Training Image set *I*
Output: Image Class C
I ← input image
L ← learning layers
N ← Network Parameter
1. for *i* = 1, …, *I*:
2. for *l*= 0, …, *L*:
3. *Pattern(i)* ← *SelectPattern(i)*
4. *Output(i)* ← *ForwardPropagation(i)*
5. *UpdateWeight* ← *Weight(Pattern(i), Output(i), N)*
6. end for
7. end for
8. *W(x)* ← Weight of each class

Return: *W(x) , Weight for target class*

4. Experimental Results and Analysis

In this study, we split this research work into two-stage where in first stage, we found the odds ratio through logistic regression and determine what effects are significant and the significant variables were further used for a comparison test to know that how much risk factors are different in CKD than NCKD.

4.1. Odds ratio and comparison

Out of 24 variables, a significant odds ratio was found for blood glucose, Serum creatinine, Sodium, Potassium, Packed Cell volume, White blood cell count, red blood cells count. The forest plot presents the results of a significant odds ratio with P-value and 95% confidence interval (CI) regarding CKD are exhibited in Table 4. The blood glucose random has 0.987 odds ratio with (0.977-0.997) and 0.012 CI and P-value respectively. Serum creatinine has 0.065 odds ratio with (0.042-0.094) and 0.001 CI and P-value respectively. Moreover, Sodium has 1.032 odds ratio with (1.017-1.047) and 0.001 CI and P-value respectively. Similarly, Potassium has 1.52 odds ratio with (1.147-2.016) and 0.004 CI and P-value, respectively. At the same time, packed cell volume has 1.059 odds ratio with (1.011-1.109) and 0.016 CI and P-value respectively. Similarly, white blood cell count has a 0.999 odds ratio with (0.969-1.1) and 0.001 CI and P-value respectively. Similarly, red blood cell count has a 4.561 odds ratio with (2.328-8.937) and 0.001 CI and P-value respectively—the significant variables T-test was used to compare CKD and NCKD subjects. The comparison of important variables show where the level of the explanatory variable increases or decreases in CKD patients than in NCKD subjects. The comparison results regarding T-value, P-value, and Mean difference are shown in Table 2. The 45.34(P<=.001) and 3.36(P<=.001) increase levels found in Blood glucose random and Serum creatinine, respectively in CKD patients.

Similarly, the decreased level found in Sodium, Potassium, packed cell volume, white blood cell count, and red blood cells count -47.02(P=.001), -.934(P=.008), -20.98(P=.001), -1868.0(P=.001), and -3.13(P=.001) respectively in CKD patients than NCKD subjects.

In stage 2, we perform prediction classifiers such as LR, LDA, and MLP for the prediction of CKD and NCKD subjects. The results of applied classifiers regarding the accuracy, sensitivity, specificity, FPR, and FNR of training and testing data are presented in Table 3.

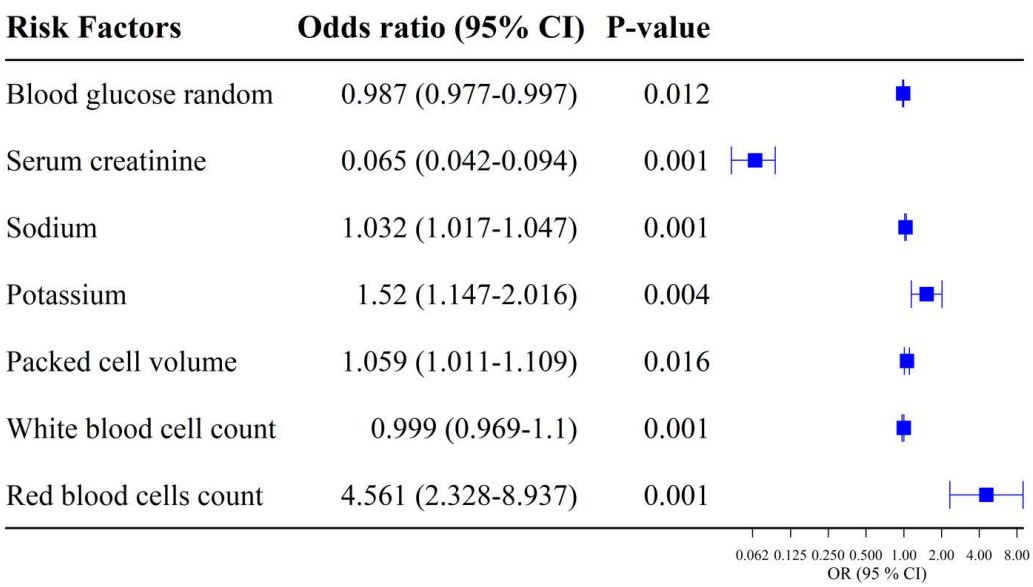

Figure 3. Forest plot presents the odds ratio.

Table 2. Comparison of CKD-NCKD through T-test.

Variables	T-value	P-value	Mean Difference
Blood glucose random	5.146	.001	45.343
Serum creatinine	6.010	.001	3.36300
Sodium	-8.606	.001	-47.025
Potassium	-2.673	.008	-.93453
Packed cell volume	-14.953	.001	-20.988
White blood cell count	-4.107	.001	-1868.000
Red blood cells count	-16.757	.001	-3.13960

Table 3. Classifiers performance comparison

	LR		LDA		MLP	
	Train	Test	Train	Test	Train	Test
Accuracy	98.5%	97.5%	96.07%	96.6%	95%	94.1%
Sensitivity	0.98	1	0.98	1	0.97	0.96
Specificity	0.99	0.93	0.92	0.91	0.91	0.90
FPR	0.017	0.0	0.017	0.0	0.029	0.038
FNR	0.009	0.06	0.07	0.08	0.080	0.093

The accuracies of LR were 98.5% and 97.5% found for train and test datasets. Moreover, accuracies of LDA were 96.07% and 96.6% found for train and test dataset. Relatedly, accuracies of MLP were 95% and 94.1% found for training and testing data, respectively. Overall, LR's accuracy was obtained higher in training and testing data than in other classifiers. The correct and incorrect prediction decision of LR, LDA, and MLP regarding training and testing data are presented in Figure 4. The LR correctly predicted the 171 observations

of CKD and 105 of NCKD out of 280 words and 4 observations were incorrectly predicted in training data. Similarly, the LDA correctly predicted the 164 observations of CKD and 105 of NCKD out of 280 observations, and 11 comments were incorrectly predicted in training data. Similarly, the MLP correctly predicted the 163 observations of CKD and 103 of NCKD out of 280 observations, and 14 words were incorrectly predicted in training data. The correct prediction rate of LR was greater than the other two models in training data. The LR correctly predicted the 75 observations of CKD and 42 of NCKD out of 120 words and 3 observations were incorrectly predicted in training data. Similarly, the LDA correctly predicted 74 observations of CKD and 42 of NCKD out of 120 observations, and four observations were incorrectly predicted in training data. Similarly, the MLP correctly predicted 74 observations of CKD and 39 of NCKD out of 120 observations, and seven observations were incorrectly predicted in training data. The correct prediction rate of LR was greater than the other two models in testing data.

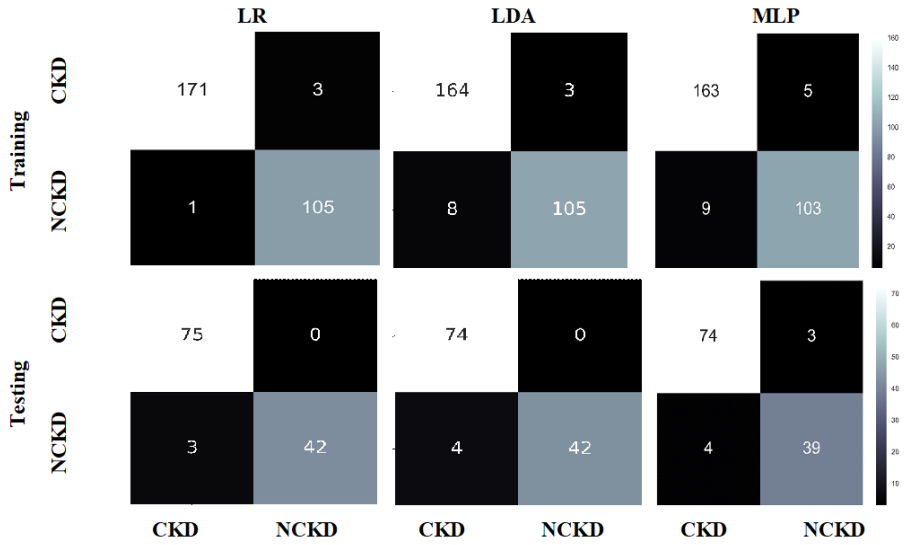

Figure 4. confusion matrices of LR, LDA, and MLP of training and testing data.

4.2. Sensitivity analysis

Sensitivity, specificity, FPR and FNR were calculated of LR, LDA, and MLP-related training and testing data to access more about the validity of the classifiers. The sensitivity is about the ability of all models to correctly identify the CKD patients if the patient belongs to the CKD group. The sensitivity values of training data were measured 0.98, 0.98, and 0.97 for LR, LDA, and MLP respectively. Similarly, sensitivity values of testing data were measured 1, 1 and 0.96 for LR, LDA, and MLP. The specificity highlights ability of the model to correctly identify the NCKD patients if the patient belongs to the NCKD group. The specificity values of training data were measured 0.99, 0.92, and 0.91 for LR, LDA, and MLP respectively. Similarly, specificity values of testing data were measured 0.93, 0.91, and 0.90 for LR, LDA, and MLP respectively. The FPR and FNR are opposite the sensitivity and specificity. The FPR means the model predicts the NCKD if it belongs to the CKD group and FNR means the model predicts the CKD if it belongs to the NCKD group. The FPR values of training data were measured 0.017, 0.017, and 0.029 for LR, LDA, and MLP. Similarly, FPR values of testing data were measured 0.0, 0.0, and 0.038 for LR, LDA, and MLP. The FNR values of training data were measured 0.009,

0.07, and 0.080 for LR, LDA, and MLP. Similarly, FNR testing data values were measured 0.06, 0.08, and 0.093 for LR, LDA, and MLP, respectively. The values of Sensitivity, Specificity, FPR, and FNR of LR, LDA, and MLP-related training and testing data are useful and valid.

4.3. ROC curve analysis

The ROC curve analysis [35] was used in this study to observe the precision and plot the discriminatory ability of correctly picking the CKD and NCKD subjects of LR, LDA, and MLP models. The ROC curve is also used to provide a high rate of sensitivity, specificity and as well as for test-retest stability. The area under the curve (AUC) is a method to diagnose classification accuracy. The value of the ROC curve between 0.70-0.80 is acceptable. Greater than 0.80 consider excellent and higher than 0.90 seems to be rarely observed. The ROC curve with AUC values of LR, LDA and MLP for training and testing data is presented in solid lines in Figure 5. The blue, green and red color solid lines present the ROC curve with AUC of LR, LDA and MLP respectively. The AUC values were 0.98, 0.96, and 0.95 of LR, LDA, and MLP respectively for training data. Similarly, AUC values were 0.98, 0.97, and 0.94 of LR, LDA, and MLP, respectively for testing data. Overall AUC value of LR was greater than the other two models.

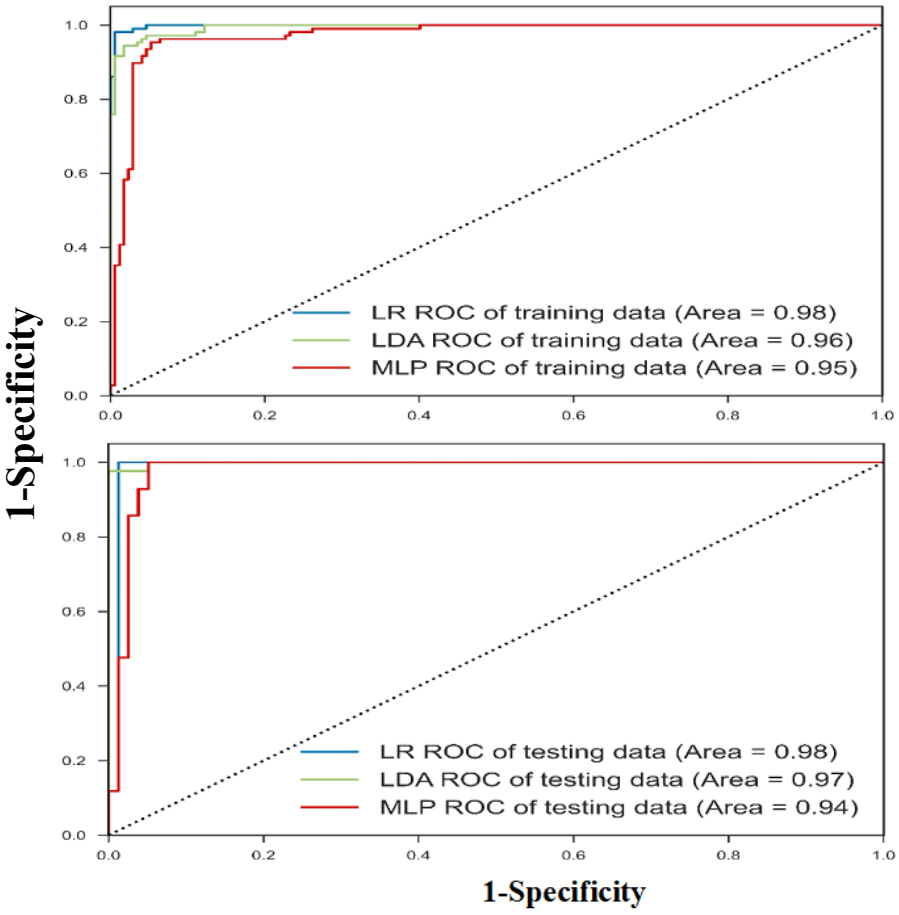

Figure 5. ROC Curve with AUC presenting the sensitivity analysis of LR, LDA, and MLP models.

4.4. Cross-validation

The 5-fold Cross-validation (CV) approach was also used in this study to prevent the overfitting of the model. The CV also estimates the prediction error of the model. The LR model was found high accurate than the other two models. The 5-fold CV approach was used with the LR model, and each fold's accuracy with respect to training and testing data is exhibited in Table 4. The accuracies of 5-fold were 0.92, 1, 0.91, 0.94, and 0.98 were found in the training dataset.data. Similarly, the accuracies of 5-fold were 1, 0.91, 0.95, 0.83, and 0.98 found in training data. The mean accuracies were 0.954 and 0.942 for training and testing data respectively.

Table 4. LR model accuracy of 5-fold CV

	Accuracy with 5-Fold CV Of LR					
	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Mean
Training Data	0.92	1	0.91	0.94	0.98	0.954
Testing Data	1	0.91	0.95	0.83	0.98	0.942

4.5. Features analysis

The features importance is the most common concerns of the classification models [36]. In this research, the LR model was found to be an efficient and reliable model for prediction. The most important features that were also extracted from the LR model are exhibited in Figure 6. In LR, serum creatinine, albumin, diabetes mellitus, red blood cells count, pus cell, hypertension and diabetics found to be the most important feature to discriminate the CKD patients from NCKD subjects. Therefore, these variables will help diagnose the procedure of CKD and NCKD subjects.

4.6. Discussion and comparisons

Practitioners and researchers want an efficient model to diagnose CKD in patients to know the critical factors involved in CKD. In this study, we split proposed model into two main stages, first , we found the odds ratio through logistic regression and determined which effects are significant. Then the important variables were further used for a comparison test to know how much and how many risk factors are different in CKD than NCKD subjects. After applying the LR, a significant odds ratio was found for Blood glucose, Serum creatinine, Sodium, Potassium, Packed Cell volume, White blood cell count, and red blood cells count. The estimated odds of blood glucose random is 0.987 with a 95% confidence interval (0.977 to 0.997); the odds are less than 1, meaning the blood glucose random is lesser in NKCD than in CKD. The odd of serum creatinine is 0.065 with a 95% confidence interval (0.042 to 0.094). The odds are less than one means the serum creatinine is less in NCKD than CKD. The odd of sodium is 1.032 with a 95% confidence interval (1.017 to 1.047). The odds are greater than one means sodium is higher in NCKD than in CKD. The odds of potassium is 1.520 with a 95% confidence interval (1.147 to 2.016). The odds are greater than 1 means potassium is higher in NCKD than CKD. The odd of packed cell volume is 1.059 with a 95% confidence interval (1.011 to 1.109) greater than 1 means packed cell volume is higher in NCKD than CKD. The white blood cell count odd is 0.999 with a 95% confidence interval (0.969 to 1.1). The odds are almost equal to 1 means white blood cell count has no effect. The odd of red blood cell count is 4.561 with a 95% confidence interval (2.328 to 8.937) the strange is greater than 1 means red blood cell count is higher in NCKD than in CKD. These are essential variables. With these variables the risk of CKD goes higher. Comparing significant odds reveals where the explanatory

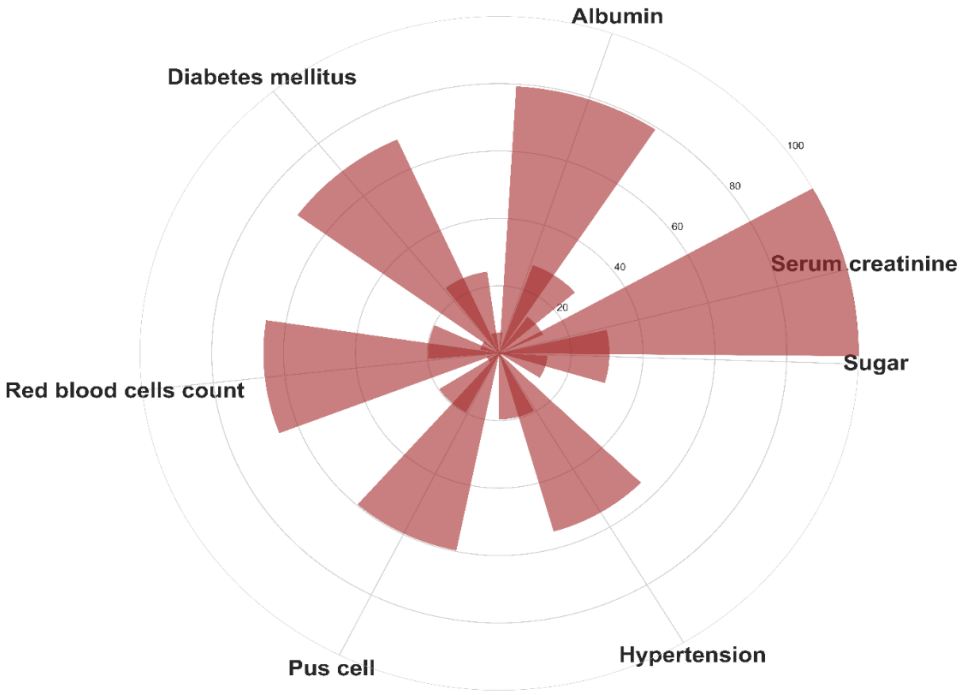

Figure 6. Circular bar graph of discriminative features selection.

variables increase or decrease in CKD patients more than in NCKD subjects. The comparison results showed increased levels found in Blood glucose random and Serum creatinine, respectively in CKD patients than in NCKD. Similarly, the decreased level was found in Sodium, Potassium, Packed cell volume, White blood cell count, and red blood cells count, respectively in CKD patients than in NCKD subjects. In stage 2, we perform prediction classifiers such as LR, LDA, and MLP to predict CKD and NCKD subjects. The train data accuracy of LR was 98.5%, and the test data accuracy was 97.5% found. Likewise, accuracies of LDA were 96.07% and 96.6% found for training and testing data. Also, accuracies of MLP were 95% and 94.1% found for training and testing data. The values of Sensitivity, Specificity, FPR, and FNR of LR, LDA, and MLP-related training and testing data are valuable and valid. The LR correctly predicted the 171 observations of CKD and 105 of NCKD out of 280 observations, and four observations were incorrectly predicted in training data. Similarly, the LR correctly predicted 75 observations of CKD and 42 of NCKD out of 120 observations, and three observations were incorrectly predicted in training data. Overall, LR's accuracy and AUC were obtained higher in training and testing data than other classifiers. The proposed study was the case-control study, and LR is often used for the case-control analysis [37, 38]. It revealed the important features of the case control study than other models [39, 40]. The target aims were accomplished from LR to find the most significant features which could classify the CKD from NCKD. So, we used the 5-fold CV approach in this study to prevent the overfitting of

the LR model. The accuracies of 5-fold were 0.92, 1, 0.91, 0.94, and 0.98 found in training data. Similarly, the accuracies of 5-fold were 1, 0.91, 0.95, 0.83, and 0.98 found in training data. The mean accuracies were 0.954 and 0.942 for training and testing data. The most important features were also extracted with the best classifier. In this research, LR model found effective and reliable model for prediction. According to LR the serum creatinine, albumin, diabetes mellitus, red blood cells count, pus cell, hypertension, and diabetics level were found to be the most important feature to discriminate CKD patients from NCKD subjects [41]. These variables will help to diagnose CKD and NCKD subjects.

4.7. Limitations and Future Work

Machine learning models require a lot of reliable and representative data to be trained and evaluated. However, it could be challenging to get comprehensive and correctly annotated datasets in the field of CKD prediction. Machine learning models' performance and generalizability may be hampered by the existence of imperfect or noisy data, inconsistent data formats, and limited sample sizes. In order to build machine learning models, it is crucial to identify useful traits in the raw data. Accurate renal prediction requires choosing qualities that capture pertinent biomarkers or indicators of kidney health, which can be difficult. A model's performance may suffer as a result of poor feature selection producing inaccurate predictions.

In the future, we may apply more feature selection strategies to extract significant information from the chronic kidney disease and design a transformative learning strategy using deep learning models. The most significant features are automatically extracted using deep learning.

5. Conclusion

In this study, we split proposed work into two-stages where in first stage, we found the odds ratio through logistic regression and comparison test to know that how much and how many risk factors are affected and different in CKD than NCKD subjects. The estimated odds revealed that blood glucose random and serum creatinine were lesser in NKCD than in CKD. Similarly, the estimated odds revealed that sodium, Potassium, packed cell volume, white blood cell count, and red blood cell count were Higher in NKCD than in CKD. The comparison results were showed increased levels found in Blood glucose random and Serum creatinine respectively in CKD patients than NCKD. Similarly, the decreased level found in Sodium, Potassium, Packed cell volume, White blood cell count, and red blood cells count respectively in CKD patients than NCKD subjects. In the second stage, we perform prediction classifiers such as LR, LDA, and MLP to predict CKD and NCKD subjects. Overall, LR's accuracy and AUC were obtained higher in training and testing data than other classifiers. So, we used 5-fold CV approach on the LR model. The mean accuracies of LR were 0.954 and 0.942 for training and testing data respectively. The LR model found a good and reliable model for prediction. According to LR the serum creatinine, Albumin, Diabetes mellitus, red blood cells count, pus cell, hypertension, and diabetics were found the most important feature to discriminate the CKD patients from NCKD subjects. These variables will help diagnose the procedure of CKD and NCKD subjects. Overall, the accuracy, 5-fold accuracy, and AUC of LR were obtained high in training and testing data than other classifiers. Therefore, the LR model was found to be a good and reliable model for prediction. The researcher or doctors should focus on blood glucose random, serum creatinine, albumin, diabetes mellitus, sodium, potassium, packed cell volume, white blood cell count, red blood cells count, pus cell, hypertension, and diabetics variables. These variables will help diagnose the CKD and NCKD subjects.

Declaration

Authors declare no conflict, and all contributed equally scientifically.

Acknowledgement

This work was supported by the research Project "Machine learning model for prediction of chronic kidney disease patients using handcrafted features for early health rehabilitation "SEED-CCIS-2022(118)", Prince Sultan University, Riyadh, Saudi Arabia". Authors are thankful for the support.

Data Availability statement

The dataset used in this study openly available from the University of California, Irvine's online machine learning repository, gets the data from chronic illness disease from a health care unit within two months. The data consist of 25 variables, where 11 are in numeric form, and 14 are in nominal form with binary and multiple labels. https://archive.ics.uci.edu/ml/index.php.

References

[1] Mujahid, M., Rehman, A., Alam, T., Alamri, F.S., Fati, S.M. and Saba, T., 2023. An Efficient Ensemble Approach for Alzheimer's Disease Detection Using an Adaptive Synthetic Technique and Deep Learning. Diagnostics, 13(15), p.2489. https://doi.org/10.3390/diagnostics13152489

[2] Muthulakshmi, I. "An Extensive Survey on Evolutionary Algorithm Based Kidney Disease Prediction." 2019 International Conference on Recent Advances in Energy-efficient Computing and Communication (ICRAECC). IEEE, 2019. DOI: 10.1109/ICRAECC43874.2019.8995152

[3] Zhang Z, Ho KM, Hong Y. Machine learning for the prediction of volume responsiveness in patients with oliguric acute kidney injury in critical care. Critical Care. 2019; 23(1):10. https://doi.org/10.1186/s13054-019-2411-z

[4] Park N, Kang E, Park M, Lee H, Kang HG, Yoon HJ, et al. Predicting acute kidney injury in cancer patients using heterogeneous and irregular data. PloS one. 2018; 13(7):21 pages. https://doi.org/10.1371/journal.pone.0199839

[5] Abunadi, I., and Senan, E. M. (2022). Multi-Method Diagnosis of Blood Microscopic Sample for Early Detection of Acute Lymphoblastic Leukemia Based on Deep Learning and Hybrid Techniques. Sensors, 22(4), 1629. https://doi.org/10.3390/s22041629

[6] National Kidney Foundation (NKF), The Facts About Chronic Kidney Disease (CKD), National Kidney Foundation, accessed on March 2019, (2019), http://www.kidney.org/kidneydisease/aboutckd.cfm

[7] Lin LY, Warren-Gash C, Smeeth L, Chen PC. Data resource profile: The National Health Insurance Research Database (NHIRD). Epidemiology and Health. 2018; 40:6 pages. https://doi.org/10.4178/epih.e2018062

[8] Mahadevan, V., 2019. Anatomy of the kidney and ureter. Surgery (Oxford), 37(7), pp.359-364. https://www.mdpi.com/2076-3417/11/18/8438

[9] Koye, D.N., Magliano, D.J., Nelson, R.G. and Pavkov, M.E., 2018. The global epidemiology of diabetes and kidney disease. Advances in chronic kidney disease, 25(2), pp.121-132. https://doi.org/10.1053/j.ackd.2017.10.011

[10] Kavoussi, N. L., Floyd, C., Abraham, A., Sui, W., Bejan, C., Capra, J. A., Hsi, R. (2022). Machine learning models to predict 24 hour urinary abnormalities for kidney stone disease. Urology, 169, 52-57. https://doi.org/10.1016/j.urology.2022.07.008

[11] Yar, H., Khan, Z. A., Ullah, F. U. M., Ullah, W., & Baik, S. W. (2023). A modified YOLOv5 architecture for efficient fire detection in smart cities. Expert Systems with Applications, 231, 120465. https://doi.org/10.1016/j.eswa.2023.120465

[12] Rahim, M. S. M., Norouzi, A., et al., (2017). 3D bones segmentation based on CT images visualization. Biomedical Research, 28(8), 3641-3644.

[13] Amin, J., Sharif, M., Raza, M., et al., (2019, April). Brain tumor classification: feature fusion. In 2019 international conference on computer and information sciences (ICCIS) (pp. 1-6). IEEE. DOI: 10.1109/ICCISci.2019.8716449

[14] Saba, T., Rehman, A., Sulong, G. (2010). An intelligent approach to image denoising. Journal Of theoretical and Applied information technology, 17(2), 32-36.

[15] Saba, T., Khan, S. U., Islam, N., Abbas, N., et al., (2019). Cloud-based decision support system for the detection and classification of malignant cells in breast cancer using breast cytology images. Microscopy research and technique, 82(6), 775-785.

[16] Dovgan, E., Gradišek, A., Luštrek, M., Uddin, M., et al., (2020). Using machine learning models to predict the initiation of renal replacement therapy among chronic kidney disease patients. Plos one, 15(6), e0233976.

[17] Hsi, R. S., Yan, P. L., Goldfarb, D. S., Egbuji, A., Si, Y., Shahinian, V., & Hollingsworth, J. M. (2021). Comparison of selective versus empiric pharmacologic preventative therapy with kidney stone recurrence. Urology, 149, 81-88. https://doi.org/10.1016/j.urology.2020.11.054

[18] S. Vijayarani and S. Dhayanand, Data Mining Classification Algorithms for Kidney Disease Prediction, International Journal on Cybernetics & Informatics (IJCI) 4. 4, (2015), 217-227.

[19] Abraham, A., Kavoussi, N. L., Sui, W., Bejan, C., Capra, J. A., & Hsi, R. (2022). Machine learning prediction of kidney stone composition using electronic health record-derived features. Journal of Endourology, 36(2), 243-250. https://doi.org/10.1089/end.2021.0211

[20] Dar, S. U., Öztürk, Ş., Korkmaz, Y., Elmas, G., Özbey, M., Güngör, A., & Çukur, T. (2022). Adaptive diffusion priors for accelerated mri reconstruction. arXiv preprint arXiv:2207.05876.

[21] Polat, H., Mehr, H. D., Cetin, A. (2017). Diagnosis of chronic kidney disease based on support vector machine by feature selection methods. Journal of medical systems, 41(4), 55

[22] Ahmad, M., Tundjungsari, V., Widianti, D., et al., (2017, November). Diagnostic decision support system of chronic kidney disease using support vector machine. In 2017 second international conference on informatics and computing (ICIC) (pp. 1-4). IEEE. DOI: 10.1109/IAC.2017.8280576

[23] Jia, C., Xu, H., & Wang, L. (2022). Robust nonlinear model predictive control for automatic train operation based on constraint tightening strategy. Asian Journal of Control, 24(1), 83-97. https://doi.org/10.1002/asjc.2419

[24] Borucka, A., & Grzelak, M. (2019). Application of logistic regression for production machinery efficiency evaluation. Applied Sciences, 9(22), 4770. https://doi.org/10.3390/app9224770

[25] Yar, H., Hussain, T., Khan, Z. A., Koundal, D., Lee, M. Y., & Baik, S. W. (2021). Vision sensor-based real-time fire detection in resource-constrained IoT environments. Computational intelligence and neuroscience, 2021. https://doi.org/10.1155/2021/5195508

[26] Jankowski, K. R., Flannelly, K. J., & Flannelly, L. T. (2018). The t-test: An influential inferential tool in chaplaincy and other healthcare research. Journal of health care chaplaincy, 24(1), 30-39.

[27] Alyami, J. (2022). Deep Learning-Based Lung Infection Detection Using Radiology Modalities and Comparisons on Benchmark Datasets in COVID-19 Pandemic. In Prognostic Models in Healthcare: AI and Statistical Approaches (pp. 483-504). Singapore: Springer Nature Singapore. DOI: 10.1109/ACCESS.2021.3098453

[28] Saba, T., Rehman, A., Mehmood, Z., Kolivand, H., Sharif, M. (2018). Image enhancement and segmentation techniques for detection of knee joint diseases: A survey. Current Medical Imaging, 14(5), 704-715.

[29] Shahzad, M. N., Ali, H., et al., (2021). Identifying patients with PTSD utilizing resting-state fMRI data and neural network approach. IEEE Access, 9, 107941-107954. DOI: 10.1109/ACCESS.2021.3098453

[30] K. B. DeSalvo, V. S. Fan, M. B. McDonell, and S. D. Fihn, ‘‘Predicting mortality and healthcare utilization with a single question,’’ Health Services Res., vol. 40, no. 4, pp. 1234–1246, Apr. 2005. https://doi.org/10.1111/j.1475-6773.2005.00404.x

[31] Özbey, M., Dar, S. U., Bedel, H. A., Dalmaz, O., Özturk, Ş., Güngör, A., & Çukur, T. (2022). Unsupervised medical image translation with adversarial diffusion models. arXiv preprint arXiv:2207.08208.

[32] Azar, A. T., Khan, Z. I., Amin, S. U., Fouad, K. M. (2023). Hybrid Global Optimization Algorithm for Feature Selection. Comput. Mater. Contin, 74, 2021-2037. DOI: 10.32604/cmc.2023.032183

[33] Mujahid, M., Rustam, F., Álvarez, R., Luis Vidal Mazón, J., Díez, I.D.L.T. and Ashraf, I., 2022. Pneumonia Classification from X-ray Images with Inception-V3 and Convolutional Neural Network. Diagnostics, 12(5), p.1280. https://doi.org/10.3390/diagnostics12051280

[34] Rehman, A., Saba, T., Mujahid, M., Alamri, F.S. and ElHakim, N., 2023. Parkinson’s Disease Detection Using Hybrid LSTM-GRU Deep Learning Model. Electronics, 12(13), p.2856. https://doi.org/10.3390/electronics12132856

[35] Yar, H., Hussain, T., Agarwal, M., Khan, Z. A., Gupta, S. K., & Baik, S. W. (2022). Optimized dual fire attention network and medium-scale fire classification benchmark. IEEE Transactions on Image Processing, 31, 6331-6343. DOI: 10.1109/TIP.2022.3207006

[36] Irshad, M., Sharif, M., Yasmin, M., Khan, M. A. (2022). Discrete light sheet microscopic segmentation of left ventricle using morphological tuning and active contours. Microscopy Research and Technique, 85(1), 308-323.

[37] Devarajan, D., Alex, D.S., Mahesh, T.R., Kumar, V.V., Aluvalu, R., Maheswari, V.U. and Shitharth, S., 2022. Cervical Cancer Diagnosis Using Intelligent Living Behavior of Artificial Jellyfish Optimized With Artificial Neural Network. IEEE Access, 10, pp.126957-126968. DOI: 10.1109/ACCESS.2022.3221451

[38] Khadidos O. Adil, Alshareef M. Abdulrhman, Manoharan Hariprasath, Khadidos O. Alaa and Shitharth Shitharth, Application of Improved Support Vector Machine for Pulmonary Syndrome Exposure with Computer Vision Measures, Current Bioinformatics 2023; vol. 18, pp. 1-13

[39] Dar, S. U., Öztürk, Ş., Özbey, M., & Çukur, T. (2023). Learning Deep MRI Reconstruction Models from Scratch in Low-Data Regimes. arXiv preprint arXiv:2301.02613.

[40] Sudheesh, R., Mujahid, M., Rustam, F., Mallampati, B., Chunduri, V., de la Torre Díez, I. and Ashraf, I., 2023. Bidirectional encoder representations from transformers and deep learning model for analyzing smartphone-related tweets. PeerJ Computer Science, 9, p.e1432. https://doi.org/10.7717/peerj-cs.1432

[41] Selvarajan, S., Manoharan, H., Iwendi, C., Alsowail, R.A. and Pandiaraj, S., 2023. A comparative recognition research on excretory organism in medical applications using artificial neural networks. Frontiers in Bioengineering and Biotechnology, 11, p.1211143. https://doi.org/10.3389/fbioe.2023.1211143