IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 3, June 2025
Impact Factor: 7.67

Data Driven Decision Making for Renal Function Assessment

Abdel Rahman Hossam Gad and Hazem M. El Bakry
Dept. of Information Systems
Faculty of Computer & Information Sciences, Mansoura University, EGYPT.

Abstract: *AI is transforming our lives. In healthcare, machine learning can predict diseases, especially in regions with limited medical resources. Chronic kidney disease (CKD) is a major global health concern. We propose the utilization of machine learning models for early detection of CKD utilization dataset, which were collected by a team of doctors in one of the top centers of treating CKD world, we aim to help the patient discover the disease early and provide suitable healthcare before reach the point of no return of the disease. We utilized three will established and common machine learning approaches (logistic regress, k nearest neighbors and support vector machine) k nearest neighbors achieved 87%, 88% ,88% 87% and 13% accuracy, precision , recall , F1 and MAE respectively outperforming both support vector machine and logistic regression*

Keywords: Chronic kidney disease

I. INTRODUCTION

Chronic kidney disease [1,2,3] is considered one of the dangerous diseases that attack the patient body as it developed over a very long-time scale which makes early detection using usual and early tools very complex. Kidney is one the most essential body organs , as it is responsible for the filtering the waste and poisonous material from the body and get rid of the excess water from the body for keeping the balance of the body. The degeneration of the function of the kidney happens over very long time and the body can adapt for the lost of the some functionality of the kidney but after a long time the body start to lose the ability to adapt and this lead to loss of the kidney and reaching the point of the total kidney failure which lead to the death of the patient without the proper care lead to imminent death, proper care includes the using of dialysis and provide kidney transplant for recovering the function of blood filters and getting ride of the excess waste and water.

Utilization machine learning [4,5,6,7] for improve human living conditions goes back to the early stages of utilizing computers. As machine learning is simply teaching the machine to act if it is like a human being, especially in the case of learning from data that is used during the training stages of the learning process. The range of applications that utilize machine learning for the improvement of human life is enormous and will need a very long discussion. We will focus on utilization machine learning approaches and other knowledge discovery approaches in healthcare and especially the utilization in the early detection of CKD

Knowledge discovery [8,9,10] is one of the most essential functions of computers in current age. As the amount of data that need to be analysis and preprocessed for extracting insight and understanding is huge and giving this job for a human is one of the hardest tasks that can be conduct by a human being. As a result, the utilization of different knowledge discovery approaches like machine learning data mining and the statistical approaches is indispensable.

Machine learning (ML) Is considered one of the widely used approaches that is under the umbrella of the AI field. It is used for many applications, like for disease predication and detection of planets disease and weather predications and many other applications, basic idea behind machine learning is simply emulating the ability of the human being of learning from the data that exist in the world around human using available sources of data like sense that he has. Simply put, giving the machine a data source and training it for extracting useful pattern of knowledge and insights.

Datamining [11,12] is simply put trying to extract insights from huge volumes of data using different approaches for preprocessing and analysis and extract knowledge from said data, the different is that It is enabling the computer to

DOI: 10.48175/IJARSCT-27520

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 3, June 2025
Impact Factor: 7.67

work with much larger amount of data compared to machine learning. It is like the process of mining rocks to extract nuggets of gold from underground using similar approaches but on different raw materials.

In this paper, we are planning to provide an approach for detection of CKD using latest approaches and providing a dataset in the process collected on the hands of top of the world physicians from top of the world centers for treating CKD. The paper is divided into introduction , related work, proposed approach of methodology, results and discussion and conclusion and future.

Related work:

One of the earliest conclusions that reader will reach when reading the literature on the CKD predication is that, it is one of the popular topics that is rich with sources of information and wealth of different approaches and application of machine learning and datamining approaches and other knowledge discovery approaches in this parts we will focus on different approaches of discovery of CKD in different stages and how to start the treatment journey early for improving the living conditions of the patients.

Islam et al [13] investigated the potential of machine learning for Chronic Kidney Disease (CKD) identification. We employ a data-driven approach that leverages predictive modeling techniques to explore the relationships between various data attributes and the target class (presence/absence of CKD).Through machine learning and predictive analytics, we construct a collection of prediction models informed by the improved feature selection capabilities provided by predictive modeling. The initial dataset included 25 variables alongside the class label. Feature selection techniques were employed to identify the most relevant subset of parameters for CKD classification, ultimately resulting in a reduction of 70% of the initial features. Within a supervised learning framework, twelve different machine learning classifiers were evaluated. The XGBoost classifier achieved the most promising performance metrics, with an accuracy of 0.983, precision of 0.98, recall of 0.98, and F1-score of 0.98.These findings suggest that advancements in machine learning, coupled with predictive modeling, offer a compelling approach for uncovering novel solutions in the realm of CKD prediction. The methodology presented here has the potential for broader application in disease prediction beyond CKD.

Arif [14] et al proposes a novel machine learning model for CKD prediction, incorporating a comprehensive pipeline that addresses common challenges in medical datasets. The model leverages various techniques to enhance its accuracy and generalizability: Preprocessing: The approach tackles missing values through iterative imputation. Additionally, a novel sequential data scaling method is introduced, combining robust scaling, z-standardization, and min-max scaling for improved data normalization. Feature Selection: The Boruta algorithm is employed for feature selection, ensuring the model focuses on the most informative features for CKD classification Model Development: The model utilizes machine learning algorithms to learn the underlying relationships between the selected features and the presence/absence of CKD The proposed model was validated on the UCI CKD dataset and achieved outstanding performance with 100% accuracy. This exceptional outcome highlights the effectiveness of the combined approach, encompassing innovative preprocessing steps, the Boruta feature selection method, the k-nearest neighbors' algorithm, and hyperparameter optimization using grid-search cross-validation.

Swain [15] et al focuses on developing a machine learning model for predicting CKD using publicly available data. To ensure the model's generalizability, a comprehensive data preprocessing pipeline was implemented: Missing Value Imputation: Missing data points were appropriately imputed to minimize their impact on the model's performance Data Balancing: The SMOTE algorithm was employed to address potential class imbalances within the dataset, ensuring the model is trained on a more balanced representation of CKD and non-CKD cases Feature Scaling: Feature scaling was performed to standardize the data and improve model convergence during training. Feature Selection: A chi-squared test was utilized to identify a minimal set of highly correlated and informative features relevant for CKD prediction. This helps reduce model complexity and potentially improves generalizability.

For model development, a stacked supervised learning approach was adopted. This involves combining multiple learning algorithms to leverage their individual strengths and potentially achieve better performance than any single model. Among the tested algorithms, Support Vector Machine (SVM) and Random Forest (RF) emerged as the most promising candidates, achieving low false-negative rates and high-test accuracy (SVM: 99.33%, RF: 98.67%). Notably,

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 3, June 2025
Impact Factor: 7.67

SVM demonstrated superior performance during 10-fold cross-validation, suggesting better generalizability to unseen data compared to RF.

Khan et al [16] compares various machine learning models for chronic kidney disease (CKD) prediction using the UCI Chronic Kidney Failure dataset. The investigation focused on accuracy, with a proposed hybrid model achieving the highest performance (99%) compared to XGBoost (98%), Random Forest (97%), Logistic Regression (94%), and AdaBoost (95%). These findings suggest the potential of ensemble methods for improved CKD prediction accuracy.

Pal et al [17] investigated machine learning models for chronic kidney disease (CKD) prediction using a UCI dataset with 25 features. Logistic Regression, Decision Tree, and Support Vector Machine classifiers were compared. The Decision Tree achieved the highest individual accuracy (95.92%). However, by employing a bagging ensemble method, the overall prediction accuracy was further improved to 97.23%. These findings highlight the potential of ensemble learning techniques to enhance CKD prediction accuracy compared to single machine learning models.

Venkatesan et al [18] explores machine learning models for chronic kidney disease (CKD) classification using a public dataset from the UCI Irvine ML Repository (400 instances). Various algorithms were evaluated, including Support Vector Machine, K-Nearest Neighbors, Random Forest, Logistic Regression, Decision Tree, and XGBoost. Performance was measured using accuracy, recall, precision, and F1-measure. XGBoost achieved the highest accuracy (98.00%), suggesting its potential for improved CKD prediction compared to other tested algorithms. This model could be valuable for policymakers to forecast CKD patterns in populations, enabling earlier detection, risk assessment, and improved resource allocation for patient-centered management.

Sawhney et al [19] proposes a Deep Neural Network (DNN) using a Multi-Layer Perceptron (MLP) classifier for chronic kidney disease (CKD) diagnosis. Trained on data from 400 patients encompassing various factors like age and blood cell counts, the model achieved perfect accuracy in classifying CKD.The research aims to introduce Deep Learning approaches for analyzing medical data and achieving accurate CKD detection. The key contribution is a DNN model reaching 100% accuracy, potentially surpassing traditional machine learning methods. The paper offers a detailed explanation of the MLP classifier built with PyTorch, highlighting the advantages of neural networks for CKD classification. These advantages include handling non-linear data complexities and facilitating independent learning through the network's layered structure.

Saif et al [20] explores three deep learning models for predicting chronic kidney disease (CKD) onset 6 or 12 months in advance. These models include a Convolutional Neural Network (CNN), a Long Short-Term Memory (LSTM) model, and a Deep Ensemble model. The Deep Ensemble approach combines the predictions of the CNN, LSTM, and LSTM-BLSTM models using majority voting. Evaluated on two public datasets, the Deep Ensemble model emerged as the superior performer, achieving impressive accuracy (0.993 for 6-month prediction and 0.992 for 12-month prediction). This suggests the potential of deep ensemble learning for earlier and potentially more accurate CKD prediction compared to individual deep learning models.

Shukla et al [21] emphasizes the importance of early chronic kidney disease (CKD) detection for improved patient outcomes. It explores the potential of Machine Learning (ML) techniques in achieving this goal. The paper examines various supervised learning algorithms, including K-Nearest Neighbors (KNN), Decision Tree, and Artificial Neural Networks (ANN), for CKD prediction using past medical data. Among these, Decision Tree emerged as the most promising, achieving an accuracy of 98.60%. Furthermore, the review discusses the role of specific attributes in CKD detection, acknowledging the influence of individual factors on disease progression. It highlights the ongoing research efforts to identify the optimal combination of ML techniques and relevant attributes for the most accurate CKD prediction.

Srivastav et al [22] investigates machine learning models for chronic kidney disease (CKD) prognosis. Given the increasing use of machine learning for disease prediction, the research addresses potential data quality issues like inaccuracies in medical records. Focusing on a UCI Machine Learning repository dataset with 25 features, the study compares Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), and Naive Bayes classifiers. Performance is evaluated using the dataset's two classes (healthy/diseased). Random Forest emerged as the most effective model, achieving an average accuracy of 89.75%.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 3, June 2025
Impact Factor: 7.67

Alsekait et al [23] proposes a novel ensemble deep learning (DL) approach for chronic kidney disease (CKD) detection. The approach incorporates multiple feature selection methods to identify the most relevant features for accurate prediction. Additionally, it integrates pre-trained DL models with a Support Vector Machine (SVM) as a meta-learner for improved performance. Experiments using a dataset of 400 patients from the UCI machine learning repository demonstrated the effectiveness of the proposed model compared to other methods. Notably, the model achieved the highest performance when using features selected through the "mutual_info_classif" method. These findings suggest that ensemble DL approaches combined with feature selection offer a promising avenue for enhanced CKD detection accuracy. This highlights the potential for improved disease prediction by leveraging both the power of deep learning and the interpretability of feature selection techniques.

Arumugham et al [24] proposes a deep learning model using Deep Neural Networks (DNNs) with an adaptive moment estimation optimizer for early chronic kidney disease (CKD) detection. Recognizing the importance of interpretability in healthcare applications, the research integrates a Local Interpretable Model-Agnostic Explainer (LIME) to explain the DNN's predictions. The model, trained on patient data using a five-layer DNN with three hidden layers, achieved an accuracy of 98.75% and an ROC AUC score of 98.86% on unseen data for CKD risk detection. LIME explanations revealed the influence of individual features on the model's predictions. This interpretability, combined with high accuracy, suggests the proposed system could be a valuable tool for medical professionals in facilitating early CKD diagnosis.

II. PROPOSED APPROACH

Dataset

Our dataset consists of up to 100 different cases, each case consists of up to three different features, the label of each data entry is simply the function of the kidney normal or impaired. Impaired means the patient suffers from CKD.

The data is divided into two equal classes. Each class has 50 cases which is fine but may introduce overfitting problems that may be needed to treat. Three features are serum creatinine, kidney size and echogenicity. They have different sets of values. All data is categorical and need to be encoded into numbers for modeling phase. Serum creatinine has a wide range of float number values, kidney size ranges from small , normal and enlarged , and the echogenicity ranged from normal, grade 1, grade 2 and grade 3.

In the next table and figure we show the distribution of the features and class in the renal function dataset.

Table 1 present the distribution of classes in the renal function dataset

Feature	Measurement	Value
Serum Creatinine	µmol/L	06 to 6.5
Kidney Size	normal	50%
	small	37%
	other	5%
Echogenicity	normal	50%
	Grade 2	29%
	other	21%
Renal Function	Normal	50%
	Impaired	50%

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 3, June 2025
Impact Factor: 7.67

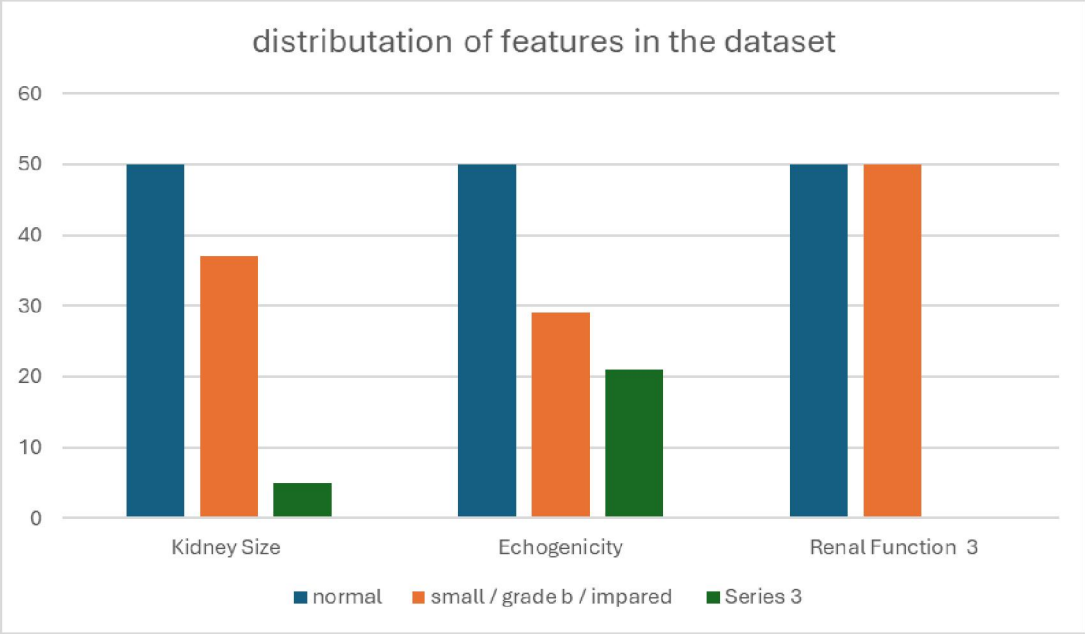

Figure 1 presents the distribution of classes in renal function dataset

Other existing datasets

Chronic Kidney Disease[25]

This dataset dates to 2015 and it was collected for two months. It contains up to 400 records and each has up to 24 different features. It suffers from some missing values which hinder and impact the performance and accuracy of the model.

Risk Factor Prediction of Chronic Kidney Disease [26]

It was collected last year in Bangladesh, and it was containing up to 200 records and has up to 28 features each. It contains no missing value which helps improve the accuracy of the model much better compared to a dataset with missing values.

Table 2 contain a comparison between our dataset and other dataset in the field

Dataset name	# Records	# Features	Missing values
Chronic Kidney Disease	400	24	yes
Risk Factor Prediction of Chronic Kidney Disease	200	28	No
Our dataset	100	3	No

Evaluation Metrics

Measuring the performance is usually done by comparing between the output of the model with the actual label in the dataset itself. This shows different aspects and view of the overall performance of the model and help us ensuring the validity and viability of the model itself [36-53].

Confusion matrix [27]

It is simply a 2 by 2 matrix that contains the number of true positive (#TP) and false negative (#FN). and the number of false positive (#FP) and the number of true negative (#TN). The first column focusses on the positive and second focuses on the negative. The first diagonal contains all true positive and negative and the second diagonal contains all

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 3, June 2025 Impact Factor: 7.67

false positive and negatives. In the following table we express the disputation of the cases between positive and negative and between true and false

Table 3 contains the distribution and number positive and negative cases in the dataset

# TP	#FN
#FP	#TN

Accuracy [28]

Is one of the most used approaches for evaluating classification or predication accuracy. It is simply calculated by calculating the total sum of all true positive and true negative on the total number of both true and false positive and negative and can be calculated from the following equation

$$Accuracy = \frac{(TP) + (TN)}{(TP) + (FP) + (TN) + (FN)} \quad (1)$$

Precision [29]

Focuses mainly on the positive cases in the classification which is simply can be calculated using by dividing True positive by the sum of the true positive and false positive and can be express in equation 2 . in some other field of science rather than computer science and machine learning it is called positive predictive.

$$Precision = \frac{(TP)}{(TP) + (FP)} \quad (2)$$

Recall [29]

It focuses mainly on the classification of positive as positive. It measures the accuracy of accurately classifying positive cases as positive and in other fields it may be called sensitivity and can be calculated using equation 3.

$$Recall = \frac{(TP)}{(TP) + (FN)} \quad (3)$$

F1-measure [29]

It simply combined precision and recall into one metric. Simply put, it is the harmonic mean of both precision and recall . it is widely used but, in some cases, it is not useful in cases of imbalanced datasets. Can be calculated from equation 4.

$$F1 = 2 \times \frac{Recall \times Precision}{Recall + Precision} \quad (4)$$

Mean absolute error [30]

One of most widely used evaluation metrics in machine learning is the average of norm 2 normalization. And it can be calculated from the following equation 5

$$MAE = \frac{1}{n}\sum_{n=1}^{n} |x_i - \hat{x}_i| \quad \frac{(TP) + (TN)}{(TP) + (FP) + (TN) + (FN)} \quad (5)$$

Where x = (1 , 2 , 3 , 4 , 5 …. n) and $\hat{x}_i$ represent the model predication.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 3, June 2025
Impact Factor: 7.67

Proposed approach

In our initial study on the renal function dataset were focus on utilizing appropriate machine learning algorithms, we did not consider deep learning approaches as they are more appropriate with huge amount of data which is not available in current dataset, we focus on using well established and common machine learning algorithms like logistic regression and k nearest neighbors and support vector machine.

Logistic Regression [31]

it is simply one of the most powerful and simplest approach for machine learning algorithms. It is under the group of supervised machine learning algorithms. It uses mathematics for the calculation of the relation between two data factors. After understanding this relation between them. It can predict the existence of one data factor given the other. It is mostly binary predicator

K Nearest Neighbors [32]

It is one of the most widely used machine learning algorithms and it fails under the group supervised machine learning algorithm. It and be used in both classification and regression. It is powerful yet very easy to implement. It uses the distance between a new point and an existing grouping of points.

Support vector machine [33]

It is a supervised machine learning algorithm that is used in classification and regression. It is simple and based on the boundaries between data points give a specific class or label. Utilizing data transformation and hyperplane to find the boundary between data points even in cases that it does not exist easily

III. RESULTS AND DISCUSSION

we applied the three machine learning algorithms without any preprocessing rather than analysis of the data and exploring the existence of the missing data points, which does not exist in our dataset as we have mentioned before. The results come as follows logistic regression achieved 80% , 86%, 75%, 80%,20% accuracy, precision , recall , F1 and MAE respectively, and K nearest neighbors achieved 87%, 88% ,88% 87% and 13% accuracy, precision , recall , F1 and MAE respectively, and finally Support vector machine achieved 80% , 86%, 75%, 80%,20% accuracy, precision , recall , F1 and MAE respectively. In the end k-neighbors achieve the highest accuracy and performance both logistic regression and support vectors machine. The results of the three models are summarized in table 4.

Table 4 showing the algorithms and their results using diverse set of metrics

Model	Accuracy	Precision	Recall	F1	MAE
Logestic regression	80%	86%	75%	80%	20%
K nearest neighbors	87%	88%	88%	87%	13%
Support vector machine	80%	86%	75%	80%	20%

In figures 2 and 3 we express the confusion matrix of the logistic regression during the training and testing phase, in figure 4 and 5 we show the results of the confusion matrix of both k nearest neighbors in both training and testing phase. And finally in figure 6 and 7 we shows the confusion matrix of training and testing of support vector machine.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 3, June 2025
Impact Factor: 7.67

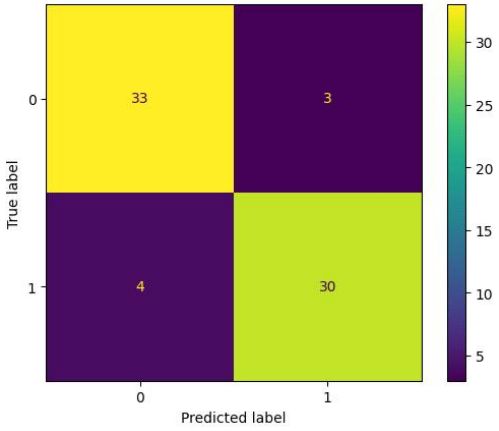

Figure 3 training phase of Logestic regression

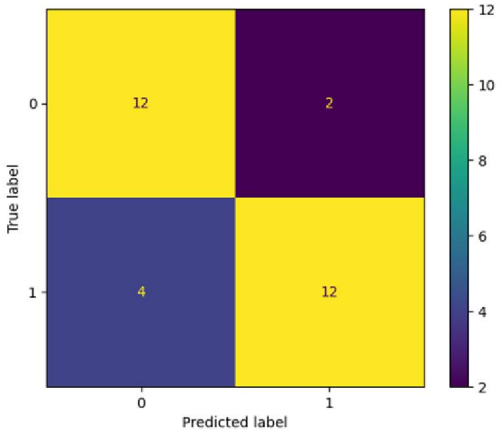

Figure 2 testing phase of Logestic regression

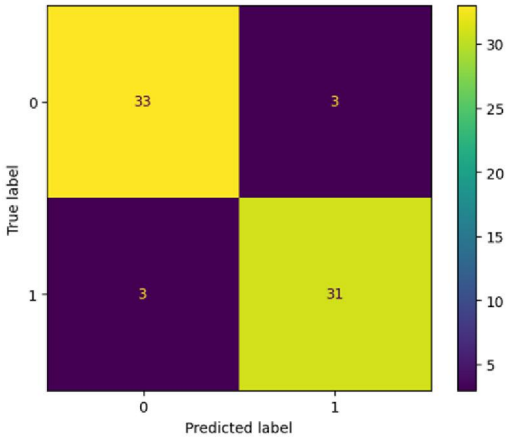

Figure 4 training phase of k-nearest neighbors

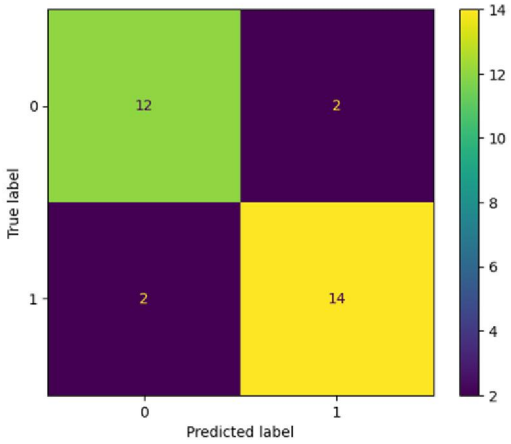

Figure 5 testing phase of k-nearest neighbors

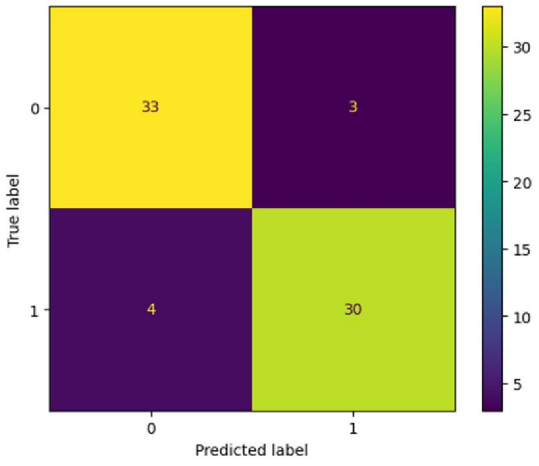

Figure 6 training phase of support vector machine

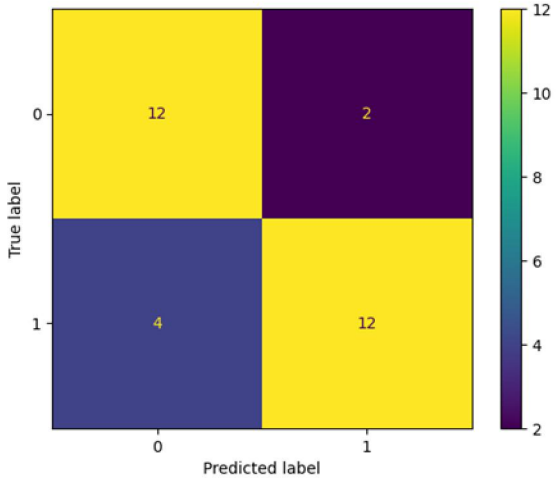

Figure 7 testing phase of support vector machine

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 3, June 2025
Impact Factor: 7.67

IV. CONCLUSION AND FUTURE WORK

We applied conventional machine learning algorithms which are suite for limited number of records and tabular data. We ignored deep learning approaches as it is limited functionality in case of limited number of records in tabular data. It records thousands or hundreds of thousands of records to reach optimal performance our proposed approach achieved good results on renal function dataset. which is promising and is the first step toward further study and understanding of the data as corner stone for building other applications based on the dataset like expert system or remote medicine applications.

We further suggest the applications of some sampling approaches for increasing the number of samples in the dataset like generative adversial network (GAN) [34] or smote [35] ,. Which utilizing the existing the distribution of data and does not suffer from over fitting which can affect other approach like random copy and pasting of existing data to increase the distribution of specific class or all classes in the dataset like in our case. We suggest also as a part of a complete set of future work increasing the number of machine learning approaches to get a wide understanding of existing datasets.

REFERENCES

[1]. Ali MM, Parveen S, Williams V, Dons R, Uwaifo GI. Cardiometabolic comorbidities and complications of obesity and chronic kidney disease (CKD). Journal of Clinical & Translational Endocrinology. 2024 Apr 2:100341.

[2]. Jabeen S, Noor S. Oxidative stress and chronic kidney disease. Fundamental Principles of Oxidative Stress in Metabolism and Reproduction. 2024 Jan 1:151-65.

[3]. Colombijn JM, Idema DL, van Beem S, Blokland AM, van der Braak K, Handoko ML, Huis LF, Kaul T, Kolagasigil-Akdemir N, Kusters MP, Meijvis SC. Representation of patients with chronic kidney disease in clinical trials of cardiovascular disease medications: a systematic review. JAMA network open. 2024 Mar 4;7(3):e240427-.

[4]. Ramírez JG, Islam MM, Even AI. Machine Learning Applications in Healthcare: Current Trends and Future Prospects. Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023. 2024 Feb 2;1(1).

[5]. Lubbad M, Karaboga D, Basturk A, Akay B, Nalbantoglu U, Paçal I. Machine learning applications in detection and diagnosis of urology cancers: a systematic literature review. Neural Computing and Applications. 2024 Apr;36(12):6355-79.

[6]. Daidone M, Ferrantelli S, Tuttolomondo A. Machine learning applications in stroke medicine: Advancements, challenges, and future prospectives. Neural Regeneration Research. 2024 Apr 1;19(4):769-73.

[7]. Bommu R. Machine Learning Applications in Cardiology: A Viable Practical Solution for Developing Countries. The Metascience. 2024 May 13;2(1):10-20.

[8]. Llatas C, Soust-Verdaguer B, Torres LC, Cagigas D. Application of Knowledge Discovery in Databases (KDD) to environmental, economic, and social indicators used in BIM workflow to support sustainable design. Journal of Building Engineering. 2024 Aug 15;91:109546.

[9]. Jha R. Development of Knowledge Discovery from Data Mining in Human Resource Application. In2024 International Conference on Advances in Modern Age Technologies for Health and Engineering Science (AMATHE) 2024 May 16 (pp. 1-5). IEEE.

[10]. Jha R. Development of Knowledge Discovery from Data Mining in Human Resource Application. In2024 International Conference on Advances in Modern Age Technologies for Health and Engineering Science (AMATHE) 2024 May 16 (pp. 1-5). IEEE.

[11]. Rastogi R, Bansal M. Diabetes prediction model using data mining techniques. Measurement: Sensors. 2023 Feb 1;25:100605.

[12]. Dol SM, Jawandhiya PM. Classification technique and its combination with clustering and association rule mining in educational data mining—A survey. Engineering Applications of Artificial Intelligence. 2023 Jun 1;122:106071.

DOI: 10.48175/IJARSCT-27520

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 3, June 2025
Impact Factor: 7.67

[13]. Islam MA, Majumder MZ, Hussein MA. Chronic kidney disease prediction based on machine learning algorithms. Journal of pathology informatics. 2023 Jan 1;14:100189.

[14]. Arif MS, Mukheimer A, Asif D. Enhancing the early detection of chronic kidney disease: a robust machine learning model. Big Data and Cognitive Computing. 2023 Aug 16;7(3):144.

[15]. Swain D, Mehta U, Bhatt A, Patel H, Patel K, Mehta D, Acharya B, Gerogiannis VC, Kanavos A, Manika S. A robust chronic kidney disease classifier using machine learning. Electronics. 2023 Jan 1;12(1):212.

[16]. Khan RH, Miah J, Rahat MA, Ahmed AH, Shahriyar MA, Lipu ER. A Comparative Analysis of Machine Learning Approaches for Chronic Kidney Disease Detection. In2023 8th International Conference on Electrical, Electronics and Information Engineering (ICEEIE) 2023 Sep 28 (pp. 1-6). IEEE.

[17]. Pal S. Chronic kidney disease prediction using machine learning techniques. Biomedical Materials & Devices. 2023 Mar;1(1):534-40.

[18]. Venkatesan VK, Ramakrishna MT, Izonin I, Tkachenko R, Havryliuk M. Efficient data preprocessing with ensemble machine learning technique for the early detection of chronic kidney disease. Applied Sciences. 2023 Feb 23;13(5):2885.

[19]. Sawhney R, Malik A, Sharma S, Narayan V. A comparative assessment of artificial intelligence models used for early prediction and evaluation of chronic kidney disease. Decision Analytics Journal. 2023 Mar 1;6:100169.

[20]. Saif D, Sarhan AM, Elshennawy NM. Deep-kidney: an effective deep learning framework for chronic kidney disease prediction. Health Information Science and Systems. 2023 Dec 1;12(1):3.

[21]. Shukla G, Dhuriya G, Pillai SK, Saini A. Chronic kidney disease prediction using machine learning algorithms and the important attributes for the detection. In2023 IEEE IAS Global Conference on Emerging Technologies (GlobConET) 2023 May 19 (pp. 1-4). IEEE.

[22]. Srivastav S, Guleria K, Sharma S. Predictive Machine Learning Approaches for Chronic Kidney Disease. In2023 4th International Conference for Emerging Technology (INCET) 2023 May 26 (pp. 1-5). IEEE.

[23]. Alsekait DM, Saleh H, Gabralla LA, Alnowaiser K, El-Sappagh S, Sahal R, El-Rashidy N. Toward comprehensive chronic kidney disease prediction based on ensemble deep learning models. Applied Sciences. 2023 Mar 20;13(6):3937.

[24]. Arumugham V, Sankaralingam BP, Jayachandran UM, Krishna KV, Sundarraj S, Mohammed M. An explainable deep learning model for prediction of early-stage chronic kidney disease. Computational Intelligence. 2023 Dec;39(6):1022-38.

[25]. Rubini,L., Soundarapandian,P., and Eswaran,P.. (2015). Chronic Kidney Disease. UCI Machine Learning Repository. https://doi.org/10.24432/C5G020.

[26]. Islam,Md. Ashiqul and Akter,Shamima. (2023). Risk Factor Prediction of Chronic Kidney Disease. UCI Machine Learning Repository. https://doi.org/10.24432/C5WP64.

[27]. Caelen O. A Bayesian interpretation of the confusion matrix. Annals of Mathematics and Artificial Intelligence. 2017 Dec;81(3):429-50.

[28]. Naidu G, Zuva T, Sibanda EM. A review of evaluation metrics in machine learning algorithms. InComputer Science On-line Conference 2023 Apr 3 (pp. 15-25). Cham: Springer International Publishing.

[29]. Foody GM. Challenges in the real world use of classification accuracy metrics: From recall and precision to the Matthews correlation coefficient. Plos one. 2023 Oct 4;18(10):e0291908.

[30]. Hodson TO. Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not, Geosci. Model Dev., 15, 5481–5487.

[31]. Wright, R. E. (1995). Logistic regression. In L. G. Grimm & P. R. Yarnold (Eds.), *Reading and understanding multivariate statistics* (pp. 217–244). American Psychological Association

[32]. Guo G, Wang H, Bell D, Bi Y, Greer K. KNN model-based approach in classification. InOn The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE: OTM Confederated International Conferences, CoopIS, DOA, and ODBASE 2003, Catania, Sicily, Italy, November 3-7, 2003. Proceedings 2003 (pp. 986-996). Springer Berlin Heidelberg.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 3, June 2025
Impact Factor: 7.67

[33]. Hearst MA, Dumais ST, Osuna E, Platt J, Scholkopf B. Support vector machines. IEEE Intelligent Systems and their applications. 1998 Jul;13(4):18-28.

[34]. Xu L, Skoularidou M, Cuesta-Infante A, Veeramachaneni K. Modeling tabular data using conditional gan. Advances in neural information processing systems. 2019;32.

[35]. Blagus R, Lusa L. SMOTE for high-dimensional class-imbalanced data. BMC bioinformatics. 2013 Dec;14:1-6.

[36]. Hazem M. El-Bakry "Fast Iris Detection for Personal Verification Using Modular Neural Networks," Proc. of the 7th Fuzzy Days International Conference, Dortmund, Germany, October 1-3, 2001, pp. 269-283.

[37]. Hazem M. El-Bakry, "Fast Virus Detection by using High Speed Time Delay Neural Networks," Journal of Computer Virology, vol.6, no.2, 2010, pp.115-122.

[38]. Hazem El-Bakry, "Face Detection Using Neural Networks and Image Decomposition," Proc. of INNS-IEEE International Joint Conference on Neural Networks, 12-17 May, 2002, Honolulu, Hawaii, USA.

[39]. Hazem M. El-Bakry, M. A. Abo-elsoud, and M. S. Kamel, "Fast Modular Neural Networks for Human Face Detection," Proc. of IEEE-INNS-ENNS International Joint Conference on Neural Networks, Como, Italy, Vol. III, pp. 320-324, 24-27 July, 2000.

[40]. Hazem El-Bakry, "Face Detection Using Neural Networks and Image Decomposition," Proc. of INNS-IEEE International Joint Conference on Neural Networks, 12-17 May, 2002, Honolulu, Hawaii, USA.

[41]. Hazem M. El-Bakry, and Nikos Mastorakis, "Realization of E-University for Distance Learning," WSEAS Transactions on Computers, vol. 8, issue 1, Jan. 2009, pp. 48-62.

[42]. Hazem El-Bakry: "Comments on Using MLP and FFT for Fast Object/Face Detection," Proc. of IEEE IJCNN'03, Portland, Oregon, pp. 1284-1288, July, 20-24, 2003.

[43]. Hazem M. El-Bakry, and Qiangfu Zhao, "Speeding-up Normalized Neural Networks For Face/Object Detection," Machine Graphics & Vision Journal (MG&V), vol. 14, No.1, 2005, pp. 29-59.

[44]. Hazem M. El-Bakry, "New Fast Time Delay Neural Networks Using Cross Correlation Performed in the Frequency Domain," Neurocomputing Journal, vol. 69, October 2006, pp. 2360-2363.

[45]. Hazem M. El-Bakry and Mohamed Hamada, "A New Implementation for High Speed Neural Networks in Frequency Space," Lecture Notes in Artificial Intelligence, Springer, KES 2008, Part I, LNAI 5177, pp. 33-40.

[46]. Hazem M. El-Bakry, and Qiangfu Zhao, "Fast Time Delay Neural Networks," International Journal of Neural Systems, vol. 15, no.6, December 2005, pp.445-455.

[47]. Hazem M. El-Bakry, "Human Iris Detection Using Fast Cooperative Modular Neural Nets," Proc. of INNS-IEEE International Joint Conference on Neural Networks, pp. 577-582, 14-19 July, 2001, Washington, DC, USA.

[48]. Hazem M. El-Bakry, "New Fast Principal Component Analysis for Face Detection," Journal of Advanced Computational Intelligence and Intelligent Informatics, vol.11, No.2, 2007, pp. 195-201.

[49]. Hazem M. El-Bakry, and Qiangfu Zhao, "Fast Normalized Neural Processors For Pattern Detection Based on Cross Correlation Implemented in the Frequency Domain," Journal of Research and Practice in Information Technology, Vol. 38, No.2, May 2006, pp. 151-170.

[50]. Menna_Elkhateeb, Abdulaziz_Shehab, and Hazem_El-bakry, "Mobile Learning System for Egyptian Higher Education Using Agile-Based Approach," Education Research International, Volume 2019, Article ID 7531980, 13 pages.

[51]. Hazem M. El-Bakry, "A Novel High Speed Neural Model for Fast Pattern Recognition," Soft Computing Journal, vol. 14, no. 6, 2010, pp. 647-666.

[52]. Hazem El-Bakry, "Fast Face Detection Using Neural Networks and Image Decomposition," Proc. of the 6th International Computer Science Conference, AMT 2001, Hong Kong, China, December 18-20, 2001, pp.205-215.

[53]. Hazem M. El-Bakry, "An Efficient Algorithm for Pattern Detection using Combined Classifiers and Data Fusion," Information Fusion Journal, vol. 11, issue 2, April 2010, pp. 133-148

DOI: 10.48175/IJARSCT-27520