Enhancing Chronic Kidney Disease Diagnosis: Scalable Machine Learning Models for Sustainable Healthcare Systems

Deepanshu[1,*], Kulvinder Singh[1], and Sanjeev Dhawan[1]

A R T I C L E I N F O

Article history:
Received: 15 July 2024
Revised: 22 November 2024
Accepted: 2 Fubruary 2025
Online: 15 May 2026

Keywords:
Health informatics
Artificial Intelligence
Predictive modeling
Chronic kidney disease
Clinical data analysis

A B S T R A C T

This study evaluates and compares many ML (machine learning) models applied to diagnose chronic kidney disease (CKD) using clinical patient records, emphasizing sustainable healthcare practices. The research investigates the application of three distinct scaling techniques: Standard Scaler, Min-Max Scaling, and Principal Component Analysis to enhance the predictive overall accuracy of these models, aiming for supporting the creation of more efficient and sustainable diagnostic solutions. The dataset, which included multiple clinical parameters, underwent thorough preprocessing, including handling missing values and applying feature scaling, to maintain data integrity and reliability. The models used in this analysis are Support Vector Classifier (SVC), Random Forest, AdaBoost Classifier, Logistic Regression, K-Nearest Neighbors, Decision Trees, a Hybrid Model of classifiers (K-Neighbors, Decision Tree, AdaBoost), and Neural Networks. The effectiveness of each model is measured by analyzing training accuracy, RMSE (Root Mean Square Error), and testing accuracy of all scaling techniques. Additionally, the study delivers a thorough analysis of the confusion matrix (CM) and classification reports, including metrics such as precision, support, F1 score, recall, and accuracy for each model and scaling technique. The results show insights into the efficacy of scaling methods in enhancing the capabilities in terms of ML algorithms for CKD diagnosis. This study demonstrates the effect of scaling methods on ML model performance, which could assist in developing more accurate and sustainable diagnostic tools for CKD. By integrating advanced ML techniques with sustainable healthcare goals, this research aims to support the broader objective of improving healthcare systems in an environmentally and socially responsible manner. To further enhance the interpretability and clinical applicability, the research also finds the capability of integrating these optimized models into real-world diagnostic systems, aiming to support clinicians in achieving timely and precise chronic kidney disease detection, ultimately contributing to better clinical results and sustainable healthcare practices.

1. INTRODUCTION

CKD is a prominent problem, characterized by a gradual deterioration of kidney function. This condition affects millions of individuals worldwide, that plays a major role in global morbidity and mortality. Its link to end-stage kidney disease (ESKD) and cardiovascular diseases further exacerbates its impact on health outcomes globally [1] [2]. The silent nature of its early stages makes CKD a particularly insidious threat, often going undetected until it is significantly advanced [3][4]. The economic burden of its treatment, especially in the form of renal replacement therapy, adds to the challenge, often making it inaccessible to a large segment of the affected population[5].

Healthcare is witnessing a growing use of AI and ML, with a specific focus on utilizing these technologies to identify and effectively manage chronic diseases in their early stages like CKD[6] [7]. Machine learning has displayed promise in utilizing vast and complex datasets to generate algorithms capable of predicting disease progression and supporting clinical decision-making. [8] [9]. The usage of ML in CKD care, from prevention and diagnosis to treatment, is particularly noteworthy. It offers potential solutions to the challenges of early detection and personalized treatment, crucial in managing a disease known for its asymptomatic progression[10] [11].

Although machine learning offers potential advantages, its integration into healthcare encounters several obstacles. These include concerns regarding the performance and methodological rigor of ML models and their comparison with traditional non-ML-based prediction tools [12][13]. Previous research has proven the efficacy of ML models in

---

[1] *Faculty of Computer Science & Engineering, University Institute of Engineering and Technology, Kurukshetra University, Kurukshetra, India.*
**Corresponding author:* Deepanshu; Email: deepanshumehla@gmail.com.

chronic disease prediction and raised questions about the quality and consistency of these predictive tools [14][15][16]. However, there is limited thorough research related to the implementation of particular machine learning models. and scaling methods in the accurate assessment of CKD using patient clinical records.

In these challenges, our work aims to fill this gap with the evaluation of various ML models, particularly DT, KNN, SVC, Random Forest, AdaBoostClassifier, Logistic Regression, Hybrid ML Models, and Neural Networks. Furthermore, the study delves into the effectiveness of three distinct scaling techniques - Standard Scaler, Min-Max Scaling, and PCA - in enhancing the models' predictive accuracy. Using a clinical dataset comprising key parameters relevant to kidney health, this research evaluates each model's performance in training accuracy, testing accuracy, RMSE, and crucial metrics like the confusion matrix and classification reports [17] [18].

Machine learning (ML) models help with the timely identification and effective management of CKD. By analysing extensive datasets of patient information, these models can identify patterns and factors that contribute to CKD progression, such as demographics, blood pressure levels, and glucose concentrations. Advanced algorithms can also predict the risk of disease onset in individuals by assimilating clinical data with lifestyle factors, potentially leading to timely interventions. Furthermore, machine learning aids physicians in customizing treatment plans by analyzing the effectiveness of various medications and treatment modalities on different patient groups. However, it is essential to address challenges like patient data confidentiality, transparency of the models, and their seamless incorporation into current healthcare infrastructures to fully leverage machine learning in combating chronic kidney disease. This paper helps within healthcare-oriented ML, with a specific focus on CKD, a disease which growing concern due to its prevalence and impact. By offering a complete analysis of various ML models and scaling methods, the present study helps to advance the effectiveness and scope of predictive analytics in CKD diagnosis, thereby contributing to early identification and effective management of this condition.

2. LITERATURE REVIEW

Sinha et al.[19] contrasted the classifiers' effectiveness in CKD prediction, which are SVM and KNN, finding KNN (78.75% accuracy) to be more effective than SVM (73.75% accuracy), as shown in Table 1. The execution time, overall precision and accuracy of the KNN and SVM classifiers are evaluated in this work for CKD prediction. The KNN classifier outperforms the SVM classifier in the experimental results.

Chen et al.[7]emphasized the importance of ML in CKD care, highlighting its role in replacement therapy planning of renal, early disease prevention, and its clinical as well as economic benefits. Their systematic review examined the diagnostic ML tools in the context of CKD with a focus on comparing ML-driven approaches to traditional non-ML techniques regarding their predictive performance, applications, and reporting standards.

Polat et al. [20] highlighted the significance of feature selection in boosting the predictive accuracy of ML algorithms used for CKD evaluation. In their study, a total of 6 classifiers – SVM, random forest, KNN, J48, preference tables, and Naïve Bayes. They concentrated on the use of the SVM classification algorithm, which combined both filter feature selection and wrapper techniques. This approach aimed to reduce the dimensionality of the dataset effectively, crucial for improving diagnostic precision. Their investigation's findings showed that the SVM classifier, after being fine-tuned with the filtered evaluator i.e. for subset and the Best First search method for selecting features, attained an impressive accuracy level of 98.5%.

Manal A Abdel-Fattah et al. [21] concentrate on enhancing the prediction of CKD by hybrid ML algorithm on platforms like Apache Spark. They employed feature selection methods, such as Relief-F and chi-squared, to identify significant features in the dataset. They used 6 classifiers, and results showed that the Support Vector Machine, Decision Tree, and Gradient-Boosted Trees classifiers, when used with selected features, achieved perfect accuracy of 100% when applied with optimally selected features. Their results showed that integrating feature selection strategies with ML algorithms on platforms such as Apache Spark can substantially enhance the performance of CKD prediction models. In their investigation, the effectiveness of ML-based approaches in healthcare, particularly in diagnosing diseases like CKD at an early stage, is critical for effective treatment and management.

3. MATERIALS AND METHODS

Here, we will discuss the methodologies used in our study, including the dataset utilized, preprocessing steps, exploratory data analysis, feature selection techniques, model training and associated evaluation metrics, visually outlined in the workflow diagram (Figure 1).

This study is conducted on UCI ML Repository Chronic Kidney Disease (CKD) Dataset, which underwent preprocessing for the data and EDA (Exploratory Data Analysis). The first data preprocessing phase involved cleaning and standardizing data, addressing incomplete or missing data entries, transforming features, and ensuring data quality for subsequent analysis. This was crucial in preparing the dataset for EDA, which provided initial insights into CKD factors and helped in understanding the distribution of the target class within the dataset which are CKD and non-CKD.

Table 1: presents a comparative summary of several CKD-related research studies

Authors	Years	Dataset	Tools	Classifiers	Results and Analysis	Future scope
Sinha et al. [19]	2015	UCI Machine learning repository	MATLAB	KNN, SVM	For KNN (Accuracy-78%, precision-85%, Recall-76%, F-measure-80%, For SVM (Accuracy-73%, precision-50%, Recall-100%, F-measure-66%)	Better solutions to the objective function can be determined through the assessment of newly developed classifiers.
Vasquez-Morales et al. [22]	2019	RIPS database of the Colombian Ministry of Health and Social Protection	Keras library and TensorFlow framework version 1.13.1	NN-CBR (Neural Network- Case Based Reasoning), Random Forest, SVM	Accuracy (Neural Network-95%, Random Forest-92%, SVM-61%)	Investigating through the use of neural network features for case representation and analyzing the effect of reducing case base size on CBR system performance
Islam M. et al.[23]	2023	UCI ML Repository	Jupyter Notebook web tool	LGBM, SVM, AdaBoost, GB, NB, Extra Tree, Hybrid, DT, XGBoost, KNN, SGB, RF, CatBoost, ANN	Highest Accuracy (Testing)-XgBoost with 99.16%, Lowest Accuracy(testing)-ANN with 60%	Different scaling techniques can be applied to the models as only PCA is applied in this paper.
Md. Mehedi Hassan et al. [24]	2023	UCI ML Repository	-	RT, RF, NN SVM, and BTM	Full Dataset-Accuracy-100%(NN), Sensitivity-100%(NN), Specificity-100% (NN, RF, SVM, RT, BTM) , Kappa-100%(NN) XGBoost Dataset-Accuracy-100% (SVM), Sensitivity - 100% (SVM), Specificity-100% (NN, RF, SVM), Kappa-100%(SVM)	Applying the developed models to datasets of other diseases. Also, developing more advanced and expert systems
Charleonnan et al. [25]	2016	UCI ML repository	-	KNN, SVM, LR, DT	Accuracy (SVM-98.3%, DT-94.8%, LR-96.55%, and KNN-98.1%.)	Investigating the usage of other ML models for CKD prediction
Debal et al.[26]	2022	The dataset is sourced from St. Paulo's Hospital.	-	RF, SVM, DT	The F1 score is highest for Random Forest (RF). Feature Selection with RFECV (SVM, RF)-99.8%. Highest accuracy (Binary Class)-99.8%	Including more clinical & demographic features in the dataset can improve the models' robustness and generalizability.
Nishat et al. [27]	2021	University of California, Irvine-learning repository	Python	CNN LR DT RF SVM NB MLP QDA	Accuracy: For LR is 98.25%, CNN is 78%, RF is 99.75%, QDA is 37.5%, MLP is 81.25%, NB is 96.5%, DT is 99% and NB is 96.5%.	Machine learning techniques enable the construction of an effective and dependable computer-aided diagnostic system.

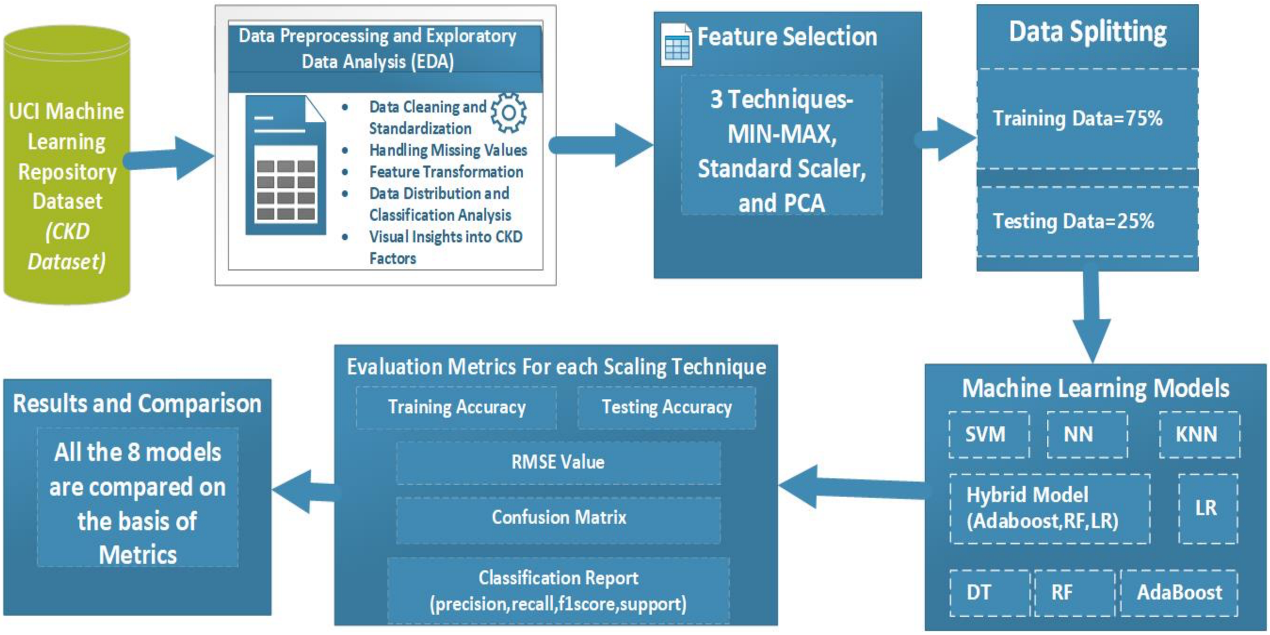
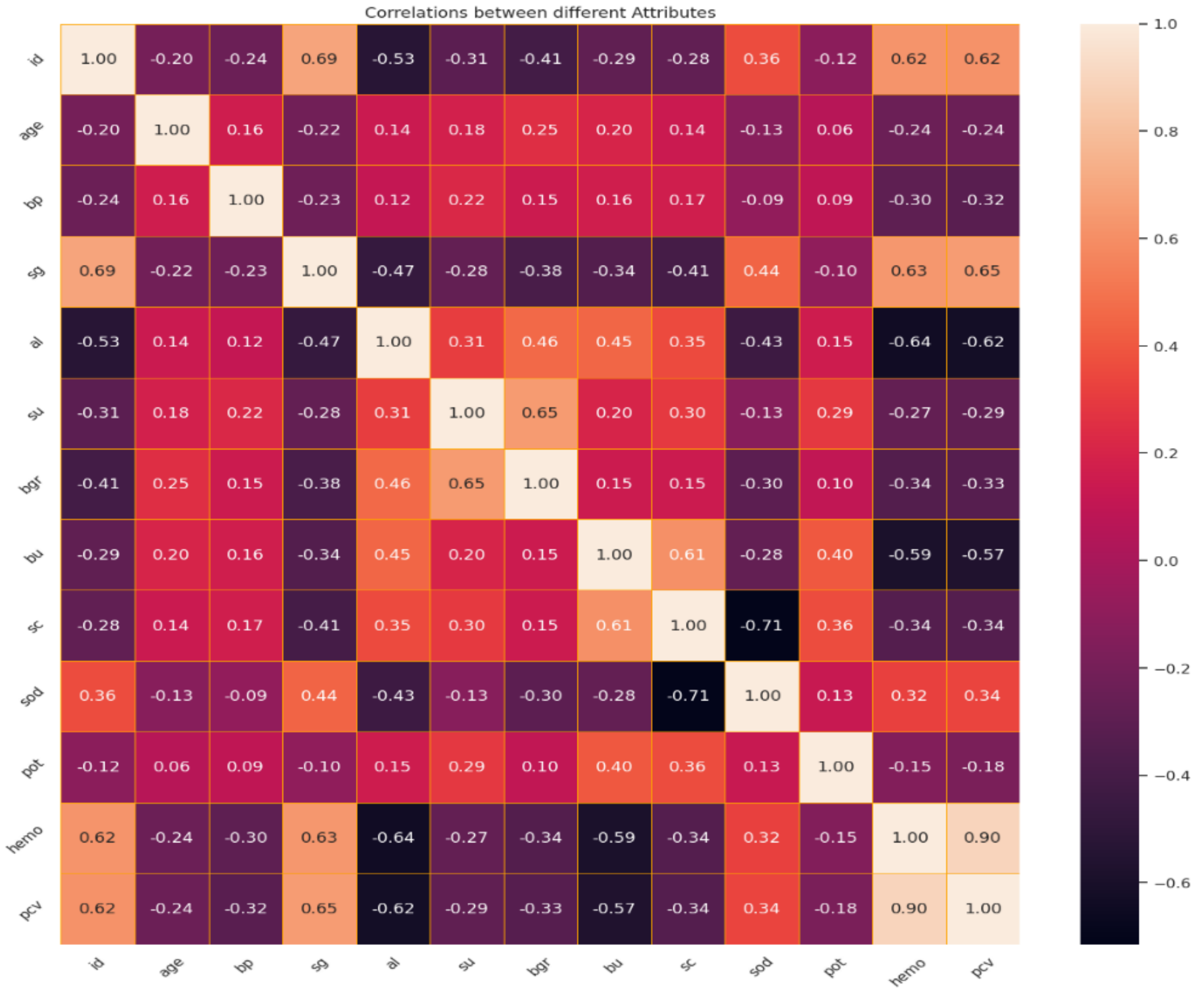


Fig. 1. The framework of the approach employed for chronic kidney disease prediction.



	id	age	bp	sg	al	su	bgr	bu	sc	sod	pot	hemo	pcv
id	1.00	-0.20	-0.24	0.69	-0.53	-0.31	-0.41	-0.29	-0.28	0.36	-0.12	0.62	0.62
age	-0.20	1.00	0.16	-0.22	0.14	0.18	0.25	0.20	0.14	-0.13	0.06	-0.24	-0.24
bp	-0.24	0.16	1.00	-0.23	0.12	0.22	0.15	0.16	0.17	-0.09	0.09	-0.30	-0.32
sg	0.69	-0.22	-0.23	1.00	-0.47	-0.28	-0.38	-0.34	-0.41	0.44	-0.10	0.63	0.65
al	-0.53	0.14	0.12	-0.47	1.00	0.31	0.46	0.45	0.35	-0.43	0.15	-0.64	-0.62
su	-0.31	0.18	0.22	-0.28	0.31	1.00	0.65	0.20	0.30	-0.13	0.29	-0.27	-0.29
bgr	-0.41	0.25	0.15	-0.38	0.46	0.65	1.00	0.15	0.15	-0.30	0.10	-0.34	-0.33
bu	-0.29	0.20	0.16	-0.34	0.45	0.20	0.15	1.00	0.61	-0.28	0.40	-0.59	-0.57
sc	-0.28	0.14	0.17	-0.41	0.35	0.30	0.15	0.61	1.00	-0.71	0.36	-0.34	-0.34
sod	0.36	-0.13	-0.09	0.44	-0.43	-0.13	-0.30	-0.28	-0.71	1.00	0.13	0.32	0.34
pot	-0.12	0.06	0.09	-0.10	0.15	0.29	0.10	0.40	0.36	0.13	1.00	-0.15	-0.18
hemo	0.62	-0.24	-0.30	0.63	-0.64	-0.27	-0.34	-0.59	-0.34	0.32	-0.15	1.00	0.90
pcv	0.62	-0.24	-0.32	0.65	-0.62	-0.29	-0.33	-0.57	-0.34	0.34	-0.18	0.90	1.00

Fig. 2. In the dataset, a heat map shows the correlation among different attributes.

For feature selection, three techniques were employed: Min-Max Scaling, Standard Scaler, and PCA. These techniques are important in normalizing the dataset and reducing dimensionality, thereby enhancing the overall efficiency of the ML based models.

The dataset is segregated into testing & training segments, from that, 75% dedicated to training purposes and 25% reserved for testing. This division was integral to confirming the performance of the models in an unbiased manner.

Several ML models are trained by the pre-processed dataset, including Neural Network, RF, KNN, SVM, AdaBoost, Decision Tree (DT), and a Hybrid ML Model (integrating AdaBoost, Random Forest, and Logistic Regression (LR)).

Evaluation metrics for each scaling technique were applied to measure the models' performance. These metrics included testing and training accuracy, RMSE value, and confusion matrix. Additionally, a classification report was generated for each model, providing F1-score, support metrics, precision, and recall.

Table 2. Shows the representation used in the dataset

Code	Description	Code	Description	Code	Description
pc	Pus Cell	bgr	Blood glucose Random	su	Sugar
htn	Hypertension	hemo	Hemoglobin	bp	Blood Pressure
rc	Blood Pressure red blood cell count	sc	Serum Creatinine	pcv	Packed cell volume
al	Albumin	wc	White Blood Cell Count	pot	Potassium
rbc	Red Blood Cells	sg	Specific Gravity	dm	Diabetes Mellitus
sod	Sodium	appet	appetite	bu	Blood Urea
Age	Age	cad	Coronary Artery Disease	ane	Anemia
pcc	Pus cell clumps	pe	Pedal Edema	class	Class

The results and comparison stage involved comparing all eight models based on the metrics to determine the optimal method for CKD prediction. That comprehensive evaluation enabled us to pinpoint the strengths and shortcomings of each model and scaling technique, facilitating an informed selection of the best-performing model.

3.1 Dataset

The CKD data obtained from the UCI Machine Learning Repository contains 400 records and 25 features, of which 11 are numerical and 14 are categorical. Key features include bp, age, sugar levels, and markers like hemoglobin and electrolytes, as shown in Table 2. Designed for classification tasks, it differentiates between both non-CKD and CKD cases. As part of the UCI ML Repository, this dataset serves as a foundational resource for developing machine learning models to detect and classify CKD at an early stage. The proportion of samples with chronic kidney disease is 62.14%, while the proportion without chronic kidney disease stands at 37.5%.

3.2 Data Preprocessing and EDA

The preprocessing and EDA stages play a huge role in organizing a dataset for later ML procedures, which makes the data understandable for model training and validation [28].

In our study, the dataset containing 400 instances with 25 attributes pertinent to kidney disease underwent preprocessing. This process involved several key steps. Firstly, we improved the readability of the dataset by renaming columns with more descriptive titles. For example, abbreviations like 'bp' and 'sg' were renamed to 'blood_pressure' and 'specific_gravity', respectively. Additionally, we addressed the issue of non-numeric data in columns such as 'red and 'white' blood cell counts by converting these values to numeric, setting non-numeric entries as 'NaN'. A significant part of our data cleaning effort involved rectifying inconsistencies and errors. This included replacing erroneous '\t?' entries with 'NaN' and standardizing entries in fields such as 'coronary_artery_disease', 'diabetes_mellitus', and 'classification' to ensure uniformity across the dataset. Lastly, we performed type conversion on the 'packed_cell_volume' column, changing it to a numeric type to better suit subsequent analytical procedures. These preprocessing steps were vital in making the dataset suitable for precise and efficient analytical evaluation.

In the EDA phase of our study, we used a variety of data visualization and analysis tools, to gain comprehensive insights into the dataset. We then conducted a correlation analysis using a heatmap to understand the interrelationships between various numerical features as shown in Fig 2. Furthermore, we create a series of grouped bar charts, allowing us to explore the relationships between key variables, including 'diabetes_mellitus', 'rbc', 'albumin', 'pus_cell_clumps', 'specific_gravity', 'pus_cell', 'hypertension', and their influence on the classification of CKD, as shown in Fig 3. Additionally, we performed statistical aggregation, computing counts, means, medians, minimums, and maximums for variables such as 'red_blood_cells' and 'albumin'. This process offered an in-depth understanding of the structure of the dataset and the distribution of key features, which is essential for effective model building and hypothesis testing in kidney disease research.

Fig. 3. Relationships among key variables and classification of CKD in the dataset.

3.3 Feature Selection Techniques

In our study, which aims to apply ML models for chronic disease prediction, feature scaling played a critical role in optimizing model performance. We employed three transformational techniques: Principal Component Analysis, Standard Scaling and Min-Max Scaling. These methods are utilized for standardizing and transforming the dataset, ensuring that our models can interpret and process the features effectively.

3.3.1 Min-Max Scaling

It is [29] is a normalization method that adjusts the range of features to a specific scale, often between zero and one. This method is especially beneficial when features differ in scale and magnitude, which guarantees that the model's performance is not overly affected with any one feature because of its scale. Min-Max Scaling maintains the distribution and relationships among the original data points.

3.3.2 Standard Scaling

Standard Scaling [30], or Z-score normalization, entails adjusting the features to the transformed values conform to a standard normal distribution with mean 0 and standard deviation 1. This method is crucial when we have features that follow a Gaussian distribution, as it aligns them to a uniform scale while retaining natural variability in the feature values.

3.3.3 Principal Component Analysis

PCA is [31] a method used for the reduction of dimensionality reorients the data into a new coordinate system where the first coordinate captures the highest variance for data projections. This technique is particularly effective in reducing the dimensionality of large datasets, enhancing model efficiency without significant loss of information. Our study utilized PCA to extract the highest pertinent features in the field of chronic disease prediction. A trade-off between dimensionality reduction and performance loss underscores the need to achieve an optimal balance among computational efficiency and performance and the retention of essential predictive features.

3.4 Machine Learning Models

3.4.1 Decision Tree (DT)

The DT technique is widely used due to its straightforwardness and ease of interpretation, which makes it a favored method for classification. It functions by continuously dividing the data according to the feature value, thus forming a decision model that resembles the structure of the tree. In a decision tree, internal nodes correspond to input features, divisions like branches reflect decision rules, and finally, leaf nodes represent the final predicted outcomes. The splitting criterion often uses Gini Impurity or Entropy. The algorithm's advantage comes from its capacity to clearly outline decision paths, allowing for a straightforward interpretation of how predictions are generated. Decision Trees are particularly useful for datasets with categorical variables [32].

3.4.2 Random Forest [RF]

This is an approach of ensemble learning [33] which works by building multiple decision trees during the training process and subsequently identifying the majority class to generate the classification result. Each tree is trained using random subsets of both data and features, a method commonly referred to as bagging (Bootstrap Aggregating). It improves the DT model by including randomness for the tree generation process, thereby cutting down the risk of overfitting and better accuracy. RF has robustness and scalability, which are suitable for many applications.

3.4.3 KNN (K-Nearest Neighbors)

KNN [34] is instance-based learning method used for both regression and classification tasks. This method assigns data points to the class that is more frequent in its k no. of closest neighbors, with a metric of distance like Manhattan distance, Euclidean distance, or others, which is based on the principle of majority voting. KNN is highly intuitive and simple, with its effectiveness heavily dependent on the selection of k and the metric for measuring distance. The algorithm's performance hinges on the selection of an appropriate k value (number of neighbors) and the chosen distance metric.

3.4.4 Support Vector Classifier [SVC]

It is a robust and flexible ML technique utilized for classification purposes[35]. SVC functions by identifying the hyperplane which most optimally partitions various classes in the feature area. SVC is recognized for its efficiency in spaces with high dimensions and its versatility in handling different types of data. For linearly inseparable data, it employs kernel functions—such as the radial basis function and polynomial kernels—to transform data so it resides within spaces of greater dimensionality, enabling linear separation. SVC is particularly effective for high-dimensional data and performs well even with limited sample sizes; however, it may demand substantial computational resources, especially when applied to large datasets.

3.4.5 AdaBoostClassifier

Adaptive boosting, or AdaBoostClassifier, is a method under the category of ensemble learning [36] which aggregates many weak models for a stronger one and a more reliable classifier. It focuses on correctly classifying the training examples that previous classifiers misclassified, thereby improving the model's overall accuracy. It enhances the efficacy of weak learning algorithms. The process minimizes an exponential loss function and produces a weighted ensemble of classifiers. AdaBoost is sensitive to noisy data but is effective for improving simple models like DTs.

3.4.6 Logistic Regression

LR is an approach applicable in solving classification challenges [37] , which are binary (using a logistic function (sigmoid), that maps predictions to probabilities between 0 and 1). It estimates the likelihood of the target variable being part of a specific class and is valuable for evaluating the effect of multiple variables on a binary result. It estimates coefficients for the target variable in a manner that maximizes the log-likelihood of the observed dataset. Its simplicity and interpretability make it a staple in various fields, particularly in medical research.

3.4.7 Hybrid ML Model (AdaBoost Classifier + Random Forest + Logistic Regression)

The Hybrid ML Model represents a method that aggregate multiple machine learning algorithms to deliver improved predictions compared to using individual models independently. The Hybrid ML Model [38], combining AdaBoost Classifier, Random Forest, and Logistic Regression, leverages the strong points of individual algorithms to make better predictive performance. This approach integrates the robust classification of Random Forest, the adaptive boosting of AdaBoost, and the probabilistic predictions of Logistic Regression. Such hybrid models are increasingly popular for their ability to deliver superior results compared to single models.

3.4.8 Neural Network

Neural Networks [39] forms a fundamental aspect of deep learning, mirroring the brain structure of humans and operational mechanisms. Comprising levels of linked nodes or neurons, these networks can learn pattern recognition and decision-making processes. Nodes apply activation functions to introduce non-linearity. These are highly flexible and powerful, able to handle complex tasks and exist in high dimensions, rendering them suitable for a wide range of applications, including both NLP and image recognition.

3.5 Evaluation Metrics and Some Key Terms

When evaluating the efficacy of our ML models in determining chronic diseases, several evaluation terms are used. These metrics help for insight into the effectiveness & the accuracy of each model. The key metrics used were Training Accuracy, Testing Accuracy, RMSE, Classification Report (including f1-score, recall, support, precision), and CM (Confusion Matrix).

3.5.1 Training and Testing Accuracy

These metrics calculate the ratio of correct predictions generated by the classifier in both the stages (training and testing) [40]. Accuracy = (Number of correct outcomes) / (Total no. of outcomes).

3.5.2 RMSE (Root Mean Square Error) Value

RMSE [41] is a common method for measuring a model's accuracy in forecasting numerical data. It plays a big role in regression analysis by measuring the discrepancy between actual/correct and predicted values.

3.5.3 Confusion Matrix (CM)

The Confusion Matrix [42] plays a vital role in classification assignments, offering a deep through evaluation of the model's outcomes for predictions. It aids in analyzing the model's effectiveness on different categories.

3.5.4 Classification Report

The Classification Report[43] encompasses key metrics such as Precision (correctness of +ve predictions), Recall (the model's capacity for detecting all pertinent examples), F1-Score (represents the average of precision and recall), and Support (the actual count of instances for every class).

3.5.5 Overfitting

This occurs when a model becomes too familiar with the training data, capturing noise and small fluctuations, that weakens its capacity to generalize.

3.5.6 Dimensionality Reduction

Dimensionality reduction, as its name implies, lessens the no. of features in a dataset while keeping the essential information intact. It improves computation speed, simplifies models, and minimizes overfitting in high-dimensional data.

3.5.7 Sensitivity

Sensitivity measures a classifier's ability to identify true positive cases correctly. For instance, in medical diagnosis, sensitivity indicates the proportion of patients with a disease that the model accurately identifies as positive, highlighting its capability to reduce false negatives.

3.5.8 k-fold cross-validation

It is a method for assessing the performance and generalizability of a machine learning algorithm by partitioning the dataset into k equal parts, or folds. Initially, the model is trained using k−1 folds, and its performance is evaluated on the leftover fold. This process is repeated k times so that every subset serves as the test set exactly once [44] – [47]. During model evaluation, we used a value of k is 10.

4. RESULT AND DISCUSSION

Three feature scaling techniques are implemented: PCA, Min-Max Scaling, and Standard Scaler, to normalize and optimize the dataset for the prediction task. The efficacy of the models was assessed (as shown in Fig 4.) based on Training Accuracy, Testing Accuracy, RMSE Value, confusion matrix, and classification report.

Min-Max Scaling Technique: Under Min-Max scaling, LR and SVC both achieved commendable Testing Accuracies of 97.9 % with low RMSE values, indicating both high accuracy and precision in predictions as shown in Fig 4. Decision Trees (DT), while perfect in Training Accuracy, showed a drop in Testing Accuracy to 91%, highlighting potential overfitting. This is further evidenced by a higher RMSE value of 0.6. The AdaBoostClassifier's Testing Accuracy of 66% with an RMSE of 1.16 is the lowest, suggesting significant improvements are needed. The Hybrid model, though, achieves a Testing Acc. of 95% & RMSE value of 0.44.

Standard Scaler Technique: When applying Standard Scaling, Random Forest (RF) stood out with perfect scores in both Training and Testing Accuracies and an RMSE of 0, indicating an excellent fit as shown in Fig 4. AdaBoost, Hybrid, and Multilayer Perceptron (MLP) all achieved a Testing accuracy of 99% with an RMSE of 0.2, suggesting these models were highly effective under this scaling technique. Also, all models performed well with this technique, demonstrating its suitability for this dataset.

PCA Technique: Principal Component Analysis (PCA) scaling showed a more varied performance across the models. Logistic Regression's performance dropped to a Testing Accuracy of 76.51% with a higher RMSE of 0.97, indicating that dimensionality reduction might have omitted relevant information for this particular model. Decision Trees and Random Forest exhibited high Training Accuracies but reduced Testing Accuracies of 81% and 82% respectively, with RMSE values above 0.8, signifying a loss of predictive precision due to PCA's dimensionality reduction as shown in Fig. 4. The Hybrid model maintained a solid Testing Accuracy of 82% with an RMSE of 0.84, indicating its robustness even after PCA application.

The classification reports further detailed the evaluation metrics for each algorithm, allowing us to examine the trade-off between the predictions of sensitivity and specificity. The classification reports indicated that models like Logistic Regression and SVC maintained high precision and recall across most feature scaling techniques, signifying their balanced classification capability. However, AdaBoost's classification report suggested a need for model refinement, as indicated by its lower metrics in the Min-Max scaling scenario.

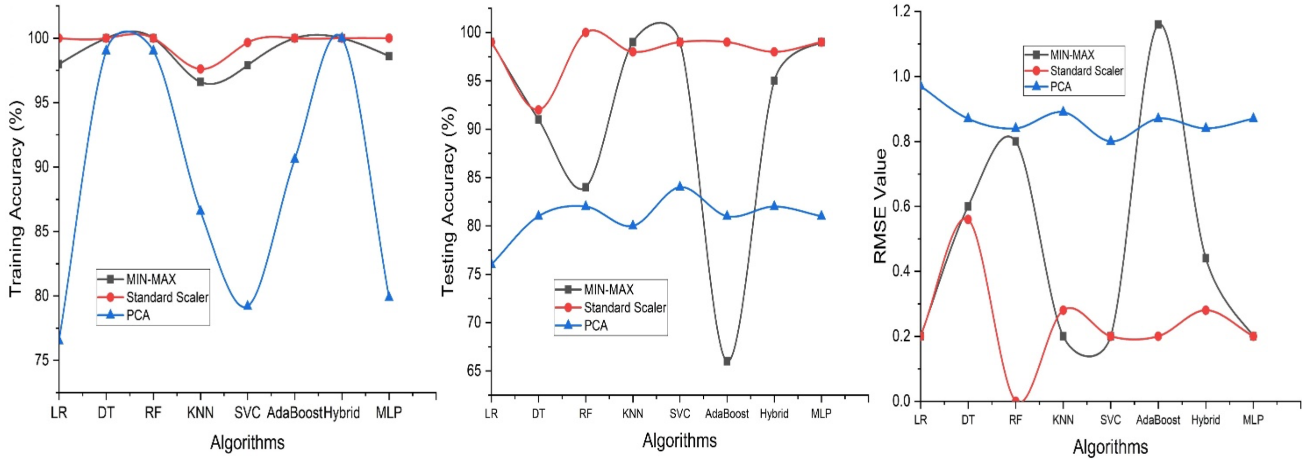


Fig. 4. The performance of ML Models across three scaling techniques (Min-Max, Standard Scaler, and PCA).

The confusion matrices provided deeper insights, revealing the distribution of TP, TN, FN, and FP. A common trend was that most models performed well in correctly identifying the non-CKD cases but had some challenges with false negatives in the CKD class which we observed in the CM for the AdaBoost, which seems critical in a medical diagnostic context where the consequences of a false negative are significant.

5. CONCLUSION AND FUTURE SCOPE

The results demonstrate that feature scaling techniques greatly influence the efficiency of ML models. The Standard Scaler technique proved to be the most effective overall, while PCA indicated potential information loss for some models. These findings highlight the importance of feature scaling and dimensionality reduction choices in the context of predictive model performance. Future work will involve investigating the cause of the high false negative rates and exploring methods to balance both specificity and sensitivity of the models further. These results show the crucial efficacy of accurate dataset preprocessing in the performance of predictive models and caution against the indiscriminate application of dimensionality reduction techniques, especially in the medical domain where the cost of prediction errors can be particularly high. Moving forward, it is imperative to address the high false negative rates observed in some models, as failing to correctly identify CKD cases can have severe implications for patient care. Therefore, future work can be focused on refining models to improve their sensitivity without compromising specificity. Techniques might involve increasing the representation of less common classes through oversampling, developing features that reveal more intricate patterns, or employing sophisticated ensemble strategies that merge several models to compensate for their individual shortcomings. By enhancing these factors, we seek to upsurge the dependability and value of ML applications in chronic disease diagnosis, ultimately contributing to better healthcare outcomes. Future studies might build upon this research by investigating the integration of renewable energy-powered healthcare infrastructure to support the deployment of ML models in resource-limited settings. Developing scalable, energy-efficient algorithms could further enhance the sustainability of diagnostic tools for chronic kidney disease (CKD). Additionally, utilizing telemedicine and mobile health technologies can enhance the availability of diagnostic services in remote and underserved communities, aligning with the goals of sustainable development. Collaborative efforts across borders to share data, resources, and expertise can foster regional and international partnerships, contributing to the collective advancement of healthcare solutions. By focusing on environmentally and socially responsible innovations, future work can ensure that the benefits of advanced machine learning techniques are equitably distributed, promoting global health and sustainability.

REFERENCES

[1] C. P. Kovesdy, “Epidemiology of chronic kidney disease: an update 2022,” Kidney Int Suppl (2011), vol. 12, no. 1, p. 7, Apr. 2022, doi: 10.1016/J.KISU.2021.11.003.

[2] D. E. Weiner et al., “Chronic Kidney Disease as a Risk Factor for Cardiovascular Disease and All-Cause Mortality: A Pooled Analysis of Community-Based Studies,” Journal of the American Society of Nephrology, vol. 15, no. 5, pp. 1307–1315, May 2004, doi: 10.1097/01.ASN.0000123691.46138.E2.

[3] A. Bellasi, L. Di Lullo, and B. Di Iorio, “chronic kidney disease: The Silent Epidemy,” J Clin Med, vol. 8, no. 11, Nov. 2019, doi: 10.3390/JCM8111795.

[4] N. Tangri, E. J. Peach, S. Franzén, S. Barone, and P. R. Kushner, “Patient Management and Clinical Outcomes Associated with a Recorded Diagnosis of Stage 3 chronic kidney disease: The REVEAL-CKD Study,” Adv Ther, vol. 40, no. 6, p. 2869, Jun. 2023, doi: 10.1007/S12325-023-02482-5.

[5] M. Tuczyńska, R. Staszewski, M. Matthews-Kozanecka, and E. Baum, “Impact of Socioeconomic Status on the Perception

of Accessibility to and Quality of Healthcare Services during the COVID-19 Pandemic among Poles—Pilot Study," Int J Environ Res Public Health, vol. 19, no. 9, May 2022, doi: 10.3390/IJERPH19095734/S1.

[6] T. Davenport and R. Kalakota, "The potential for artificial intelligence in healthcare," Future Healthc J, vol. 6, no. 2, p. 94, Jun. 2019, doi: 10.7861/FUTUREHOSP.6-2-94.

[7] S. Kumar, M. K. Chaube, S. H. Alsamhi, S. K. Gupta, M. Guizani, R. Gravina, G. Fortino, "A Novel Multimodal Fusion Framework for Early Diagnosis and Accurate Classification of COVID-19 Patients Using X-ray Images and Speech Signal Processing Techniques," Computer Methods and Programs in Biomedicine, Elsevier, 226 (2022), 107109, pp: 1-13, https://doi.org/10.1016/j.cmpb.2022.107109.

[8] H. Habehh and S. Gohel, "Machine Learning in Healthcare," Curr Genomics, vol. 22, no. 4, p. 291, Dec. 2021, doi: 10.2174/1389202922666210705124359.

[9] S. Kumar, R. Nagar, S. Bhatnagar, R. Vaddi, S. K. Gupta, M. Rashid, A. K. Bashir, T. Alkhalifah, "Chest X Ray and Cough Sample based Deep Learning Framework for Accurate Diagnosis of COVID-19", Computers and Electrical Engineering, Elsevier, 103 (2022), 108391, https://doi.org/10.1016/j.compeleceng.2022.108391.

[10] L. Chan, A. Vaid, and G. N. Nadkarni, "Applications of machine learning methods in kidney disease: Hope or Hype?," Curr Opin Nephrol Hypertens, vol. 29, no. 3, p. 319, May 2020, doi: 10.1097/MNH.0000000000000604.

[11] A. Rashid, S. K. Gupta, Z. Khanam, M. Rashid, S. S. Alshamrani, A. S. AlGhamdi, "A Novel Approach for Securing Data against Adversary Attacks in UAV Embedded HetNet using Identity Based Authentication Scheme", IET Intelligent Transport Systems, Wiley, 1-19, 2022, https://doi.org/10.1049/itr2.12271

[12] S. Uddin, A. Khan, M. E. Hossain, and M. A. Moni, "Comparing different supervised machine learning algorithms for disease prediction," BMC Med Inform Decis Mak, vol. 19, no. 1, pp. 1–16, Dec. 2019, doi: 10.1186/S12911-019-1004-8/FIGURES/12.

[13] S. Kumar, S. K. Gupta, V. Kumar, M. Kumar, M. K. Chaube, N. S. Naik, "Ensemble Multimodal Deep Learning for Early Diagnosis and Accurate Classification of COVID-19", Computers and Electrical Engineering, Elsevier, 103 (2022), 108396, pp: 1-18, https://doi.org/10.1016/j.compeleceng.2022.108396.

[14] A. Aggarwal, M. Chakradar, M. S. Bhatia, M. Kumar, T. Stephan, S. K. Gupta, S. H. Alsamhi, H. AL-Dois, "COVID-19 Risk Prediction for Diabetic Patients Using Fuzzy Inference System and Machine Learning Approaches", Journal of Healthcare Engineering, Hindawi, Vol. 2022, ID 4096950, pp: 1-10, https://doi.org/10.1155/2022/4096950

[15] Deepanshu, K. Singh, and S. Dhawan, "Diagnosing Multiple Chronic Diseases based on Machine Learning Techniques: Review, Challenges and Futuristic Approach," 2022 International Conference on Communication, Computing and Internet of Things, IC3IoT 2022 - Proceedings, 2022, doi: 10.1109/IC3IOT53935.2022.9767930.

[16] Vinayak Sharma, Nillmani, Sachin Kumar Gupta, Kaushal Kumar Shukla, "Deep learning models for tuberculosis detection and infected region visualization in chest X-ray images", Intelligent Medicine, Elsevier, 2023, https://doi.org/10.1016/j.imed.2023.06.001.

[17] L. Rubini, P. Soundarapandian, and P. Eswaran, "Chronic_Kidney_Disease - UCI Machine Learning Repository." Accessed: Dec. 27, 2023. [Online]. Available: https://archive.ics.uci.edu/dataset/336/chronic+kidney+disease

[18] A. A. Qureshi, M. Ahmad, S. Ullah, M. N. Yasir, F. Rustam, and I. Ashraf, "Performance evaluation of machine learning models on large dataset of android applications reviews," Multimed Tools Appl, vol. 82, no. 24, pp. 37197–37219, Oct. 2023, doi: 10.1007/S11042-023-14713-6/TABLES/6.

[19] Parul Sinha and Poonam Sinha, "Comparative Study of Chronic Kidney Disease Prediction using KNN and SVM," International Journal of Engineering Research and, vol. V4, no. 12, Dec. 2015, doi: 10.17577/IJERTV4IS120622.

[20] H. Polat, H. Danaei Mehr, and A. Cetin, "Diagnosis of Chronic Kidney Disease Based on Support Vector Machine by Feature Selection Methods," J Med Syst, vol. 41, no. 4, Apr. 2017, doi: 10.1007/S10916-017-0703-X.

[21] M. A. Abdel-Fattah, N. A. Othman, and N. Goher, "Predicting Chronic Kidney Disease Using Hybrid Machine Learning Based on Apache Spark," Comput Intell Neurosci, vol. 2022, 2022, doi: 10.1155/2022/9898831.

[22] G. R. Vasquez-Morales, S. M. Martinez-Monterrubio, P. Moreno-Ger, and J. A. Recio-Garcia, "Explainable Prediction of Chronic Renal Disease in the Colombian Population Using Neural Networks and Case-Based Reasoning," IEEE Access, vol. 7, pp. 152900–152910, 2019, doi: 10.1109/ACCESS.2019.2948430.

[23] M. A. Islam, M. Z. H. Majumder, and M. A. Hussein, "Chronic kidney disease prediction based on machine learning algorithms," J Pathol Inform, vol. 14, p. 100189, Jan. 2023, doi: 10.1016/J.JPI.2023.100189.

[24] M. Mahedi Hassan et al., "A Comparative Study, Prediction and Development of Chronic Kidney Disease Using Machine Learning on Patients Clinical Records," Human-Centric Intelligent Systems 2023 3:2, vol. 3, no. 2, pp. 92–104, Feb. 2023, doi: 10.1007/S44230-023-00017-3.

[25] A. Charleonnan, T. Fufaung, T. Niyomwong, W. Chokchueypattanakit, S. Suwannawach, and N. Ninchawee, "Predictive analytics for chronic kidney disease using machine learning techniques," 2016 Management and Innovation Technology International Conference, MITiCON 2016, pp. MIT80–MIT83, Sep. 2017, doi: 10.1109/MITICON.2016.8025242.

[26] D. A. Debal and T. M. Sitote, "Chronic kidney disease prediction using machine learning techniques," J Big Data, vol. 9, no. 1, pp. 1–19, Dec. 2022, doi: 10.1186/S40537-022-00657-5/FIGURES/8.

[27] M. M. Nishat et al., "A Comprehensive Analysis on Detecting Chronic Kidney Disease by Employing Machine Learning Algorithms," EAI Endorsed Trans Pervasive Health Technol, vol. 7, no. 29, pp. e1–e1, Aug. 2021, doi: 10.4108/EAI.13-8-2021.170671.

[28] K. Mahalakshmi and P. Sujatha, "The Role of Exploratory Data Analysis and Pre-processing in the Machine Learning Predictive Model for Heart Disease," Proceedings of the 2nd IEEE International Conference on Advances in Computing, Communication and Applied Informatics, ACCAI 2023, 2023, doi: 10.1109/ACCAI58221.2023.10199714.

[29] C. Buchheim and J. Kurtz, "Min-max-min robustness: a new

approach to combinatorial optimization under uncertainty based on multiple solutions," Electron Notes Discrete Math, vol. 52, pp. 45–52, Jun. 2016, doi: 10.1016/J.ENDM.2016.03.007.

[30] M. S. Gal and D. L. Rubinfeld, "Data standardization," New York University Law Review, vol. 94, no. 4, pp. 737–770, Oct. 2019, doi: 10.2139/SSRN.3326377.

[31] I. T. Jollife and J. Cadima, "Principal component analysis: a review and recent developments," Philos Trans A Math Phys Eng Sci, vol. 374, no. 2065, Apr. 2016, doi: 10.1098/RSTA.2015.0202.

[32] W. Y. Loh, "Classification and regression trees," Wiley Interdiscip Rev Data Min Knowl Discov, vol. 1, no. 1, pp. 14–23, Jan. 2011, doi: 10.1002/WIDM.8.

[33] L. Breiman, "Random forests," Mach Learn, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324/METRICS.

[34] R. O. Duda, P. E. Hart, and D. G. Stork, "Chapter 9 - Algorithm-Independent Machine Learning," Pattern Classification, p. 680, 2001, Accessed: Jan. 12, 2024. [Online]. Available: https://www.wiley.com/en-ca/Pattern+Classification%2C+2nd+Edition-p-9780471056690

[35] C. Cortes, V. Vapnik, and L. Saitta, "Support-vector networks," Machine Learning 1995 20:3, vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: 10.1007/BF00994018.

[36] Y. Freund and R. E. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting," J Comput Syst Sci, vol. 55, no. 1, pp. 119–139, Aug. 1997, doi: 10.1006/JCSS.1997.1504.

[37] D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant, "Applied Logistic Regression: Third Edition," Applied Logistic Regression: Third Edition, pp. 1–510, Aug. 2013, doi: 10.1002/9781118548387.

[38] T. G. Dietterich, "Ensemble methods in machine learning," Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 1857 LNCS, pp. 1–15, 2000, doi: 10.1007/3-540-45014-9_1/COVER.

[39] S. Mondal, M. Shafi, S. Gupta, S. K. Gupta, "Blockchain based Secure Architecture for Electronic Healthcare Record Management," GMSARN International Journal, 16 (4), 413-426, 2022.

[40] C. Mishra and D. L. Gupta, "Deep Machine Learning and Neural Networks: An Overview," IAES International Journal of Artificial Intelligence (IJ-AI), vol. 6, no. 2, p. 66, Jun. 2017, doi: 10.11591/IJAI.V6.I2.PP66-73.

[41] S. A. Hicks et al., "On evaluation metrics for medical applications of artificial intelligence," Sci Rep, vol. 12, no. 1, Dec. 2022, doi: 10.1038/S41598-022-09954-8.

[42] D. S. K. Karunasingha, "Root mean square error or mean absolute error? Use their ratio as well," Inf Sci (N Y), vol. 585, pp. 609–629, Mar. 2022, doi: 10.1016/J.INS.2021.11.036.

[43] G. M. Ifraz, M. H. Rashid, T. Tazin, S. Bourouis, and M. M. Khan, "Comparative Analysis for Prediction of Kidney Disease Using Intelligent Machine Learning Methods," Comput Math Methods Med, vol. 2021, 2021, doi: 10.1155/2021/6141470.

[44] Ž. Vujović, "Classification Model Evaluation Metrics," International Journal of Advanced Computer Science and Applications, vol. 12, no. 6, pp. 599–606, 2021.

[45] T. T. Wong, "Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation," Pattern Recognit, vol. 48, no. 9, pp. 2839–2846, Sep. 2015.

[46] KS, A. K., Sarita, K., Kumar, S., Saket, R. K., & Swami, A., "Machine Learning-based Approach for Prevention of COVID-19 using Steam Vaporizer," GMSARN International Journal 16 (2022) 399-404, Aug. 2021

[47] Thanh, B. D., "Rice Leaf Diseases Detection Using Deep Learning Ensemble Model," GMSARN International Journal 18 (2024) 308-314, April 2023.