Journal of Scientific & Industrial Research
Vol. 84, November 2025, pp. 1179-1191
DOI: 10.56042/jsir.v84i11.11248

An Efficient Machine Learning Technique to Predict Chronic Kidney Disease (CKD)

Monisha Dey[1]* & A H M Kamal[2]

[1]Department of Computer Science and Engineering, The Millennium University, Rajarbagh, Dhaka 1217, Bangladesh
[2]Department of Computer Science and Engineering, JatiyaKabiKazi Nazrul Islam University, Trishal, Mymensingh 2224, Bangladesh

Received 19 May 2024; revised 25 July 2024; accepted 20 November 2025

Chronic Kidney Disease (CKD) is a pressing global health issue that affects millions of individuals and requires early prediction to reduce severe complications and improve clinical outcomes. In this study, an efficient machine learning framework is proposed to predict CKD with high accuracy and generalization ability. Feature significance was analyzed using three statistical measures—heatmap correlation analysis, information gain, and standard deviation—calculated for average, minimum, and maximum values. To evaluate predictive performance, Logistic Regression (LR), Support Vector Machine (SVM), and Multi-Layer Perceptron (MLP) classifiers were employed. Experimental findings reveal that the average feature scores deliver the highest classification accuracy (100%), while the minimum values yield reduced model complexity without sacrificing performance. Specifically, by eliminating nine non-critical features (specific gravity, packed cell volume, hemoglobin, red blood cell count, bacteria, coronary artery disease, pus cell clumps, anemia, and pedal edema), Logistic Regression achieved 100% accuracy. Likewise, the maximum-value-based evaluation reported an accuracy of 96.75% when one redundant feature was removed. These results demonstrate the effectiveness of selective feature elimination in minimizing computational load while maintaining robust prediction capability. The proposed methodology enhances early CKD detection, supports timely medical interventions, and offers a cost-efficient diagnostic tool. The novelty of this work lies in the integration of three feature importance measures to optimize model performance, thereby contributing a mathematically balanced and clinically significant framework for CKD prediction.

Keywords: Artificial intelligence, Clinical decision support, Feature selection, Healthcare analytics, Logistic regression

Introduction

Chronic Kidney Disease (CKD) is a leading global cause of mortality and disability, posing a serious health concern. It gradually impairs kidney function, most often due to high blood pressure or diabetes, though genetic factors may also contribute. Common symptoms include fatigue, nausea, urinary issues, and loss of appetite.[1] In advanced stages, dialysis is often required, but it is not a cure—only a temporary aid. Early diagnosis is therefore vital to slow disease progression and improve quality of life. Machine learning (ML) models are increasingly applied for CKD prediction due to their automation and growing acceptance among healthcare professionals. However, their reliability depends on accurate detection. Several state-of-the-art studies have addressed this. Almansour *et al.*[2] conducted a comparative study using neural networks and support vector machines for chronic kidney disease prediction, demonstrating the superior performance of AI-based models in medical diagnostics. Ampavathi and Saradhi[3] proposed a hybrid deep learning–based framework capable of predicting multiple diseases with high accuracy, emphasizing the adaptability of deep models in complex medical diagnosis. Bai *et al.*[4] applied machine learning techniques to forecast end-stage kidney disease among CKD patients, showing strong predictive capability for disease progression. Battineni *et al.*[5] reviewed various machine learning models for chronic disease diagnosis, noting that SVM, logistic regression, and clustering methods achieved up to about 90% accuracy in different studies. Ebiaredoh-Mienye *et al.*[6] developed a machine-learning model using filter-based feature selection to enhance chronic kidney disease prediction, achieving an accuracy of approximately 98% and demonstrating improved diagnostic performance. Fatema *et al.*[7] conducted an analytical study on chronic kidney disease and associated lifestyle risk factors among patients from multiple hospitals in Dhaka city, Bangladesh—highlighting

---

*Author for Correspondence
E-mail: badhondey1998@gmail.com

significant links between dietary patterns, hypertension, diabetes, and CKD prevalence, and emphasizing the role of lifestyle management in disease prevention. Gudeti *et al.*[8] proposed a novel machine-learning approach to predict chronic kidney disease—demonstrating improved accuracy and reliability compared to conventional models, and highlighting the potential of AI in early CKD detection. Halder *et al.* (2024)[9] developed ML-CKDP, a machine-learning-based framework for chronic kidney disease prediction integrated with a smart web application—enabling real-time risk assessment and supporting early clinical decision-making. Hema & Prasanna[10] analyzed the impact of different feature-selection techniques on the early prediction of chronic kidney disease—showing that proper feature selection significantly improves model accuracy and efficiency in CKD diagnosis. Islam *et al.*[11] applied machine-learning algorithms to predict chronic kidney disease risk factors—demonstrating effective identification of key contributors to CKD development. Islam & Islam[12] developed machine-learning models for chronic kidney disease prediction—analyzing multiple algorithms and clinical features, achieving high accuracy, and demonstrating the potential of AI to support early detection and timely intervention in CKD management. After preprocessing the CKD dataset and applying principal component analysis to identify dominant features, twelve models—AdaBoost, Decision Tree, XGBoost, CatBoost, KNN, Random Forest, Gradient Boosting, Stochastic Gradient Boosting, LightGBM, Extra Tree,[13,14] SVM, ANN, and Hybrid ML—were tested. Hybrid learning and XGBoost performed best, with XGBoost slightly ahead, achieving 98.33% accuracy and a 98% F-score. Sourabh Pal[15] applied machine-learning techniques for chronic kidney disease prediction—evaluating multiple models and feature sets, and showing improved accuracy and reliability for early CKD detection. Polat *et al.*[16] developed a support vector machine–based model for chronic kidney disease diagnosis, using feature-selection methods to enhance accuracy and improve predictive performance. Sharma *et al.*[17] proposed a drug-sensitivity prediction framework using ensemble and multi-task learning—demonstrating improved prediction accuracy and potential applications in personalized medicine.

However, existing schemes often suffer from excessive model complexity, low accuracy, traditional application methods, or weak feature reduction performance.[18] To address this, the study applied a novel feature reduction technique using heatmap, information gain, and standard deviation metrics. For each feature, maximum, minimum, and average values of these statistics were computed, ranked, and iteratively removed until performance declined.[19] This method effectively eliminated features weakly linked to their classes[20], reducing system complexity while improving accuracy.[21] The resulting compact and informative feature set enables more confident clinical analysis and advances state-of-the-art CKD prediction.

Several gaps in the existing research have been highlighted by the review. Although numerous studies have focused on improving prediction accuracy, limited attention has been given to computational efficiency and the practical applicability of these models in real-world clinical settings.[22] Furthermore, the integration of innovative feature importance metrics has remained largely underexplored.

To address these gaps, a comprehensive approach has been employed in this study, wherein heatmap analysis, information gain, and standard deviation have been combined to assess feature importance.[23] By applying these metrics to evaluate machine learning models such as Logistic Regression, SVM, and MLP, enhancements in CKD prediction accuracy and model efficiency are aimed to be achieved.

In conclusion, although significant progress has been observed in the application of ML for CKD prediction, a novel approach has been proposed in this study—one that integrates innovative metrics for evaluating feature importance. This approach is expected to lead to more accurate and efficient predictive models. By critically analyzing previous studies, identifying research gaps, and clarifying the contributions of this work, the editor’s requirements for a more analytical and comprehensive literature review have been fulfilled.

Materials and Methods

CKD Dataset Description `

The Chronic Kidney Disease dataset comes from the UCI Machine Learning Repository, and serves as a fundamental component of this research. This dataset encompasses a comprehensive collection of health-related attributes for individuals, providing valuable insights into the factors contributing the chronic kidney elements. This dataset contains 400 individual records of patients with 24 independent variables and one class variable. The class variable states whether a person is suffering from a CKD problem or not. The dataset contains 250 records for CKD patients and 150 records for normal persons.

Proposed Methodology

In the proposed methodology, a systematic process was employed to enhance ML models for CKD prediction. The raw dataset was pre-processed using mean imputation for missing values and label encoding for categorical variables.[24] Feature selection, based on heatmap analysis, information gain, and standard deviation², identified the most relevant variables, improving interpretability and reducing dimensionality. Selected features were normalized via Min-Max scaling to ensure consistency. These processed features were then applied across various ML algorithms, forming a rigorous framework aimed at maximizing both prediction accuracy and interpretability (Fig. 1)

1) **Prepossessing the CKD Dataset:** Before the dataset was applied to the proposed model, a comprehensive data preprocessing task was conducted to enhance both the quality and reliability of the CKD data and to make them usable for analysis. In the figure, the massive presence of null values can be visualized, indicating the necessity of preprocessing. These preprocessing tasks are described as follows:

a. Dataset Collection and Preprocessing: A CKD dataset is collected from a publicly available source. The dataset is then preprocessed and cleaned to handle

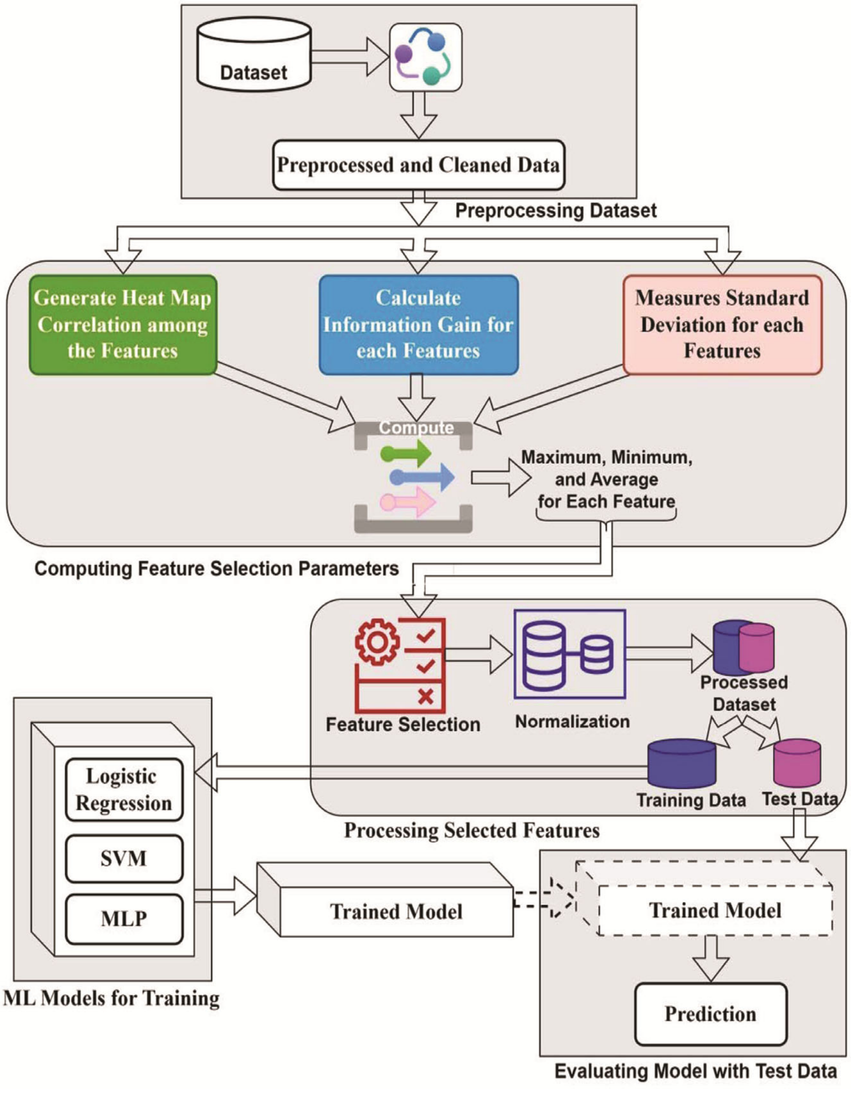

Fig. 1 — Proposed methodology: The functioning flow of the model for training a dataset and testing a sample in it

missing values, remove inconsistencies, and encode categorical variables into numerical formats. Techniques such as label encoding or one-hot encoding are applied where appropriate. The data is also checked for outliers and irrelevant instances, ensuring that it is clean and ready for analysis.

b. Heatmap Correlation Analysis: A correlation matrix heatmap is generated to identify relationships between pairs of features. This step helps in visualizing multicollinearity. Features that are highly correlated with each other but not with the target variable are marked for potential removal. Python libraries such as Pandas and Seaborn are used for this visualization.

c. Information Gain Calculation: Each feature's information gain is calculated to assess its contribution to reducing entropy in classification. This metric helps determine the relevance of a feature in distinguishing between CKD and non-CKD cases. Features with higher information gain values are given higher priority in selection. Scikit-learn or custom entropy-based functions are employed to perform this analysis.

d. Standard Deviation Measurement: The standard deviation of each feature is calculated to assess its variability across the dataset. Features with very low variance are often uninformative and may be excluded. This step ensures that the retained features have meaningful variability. This is implemented using NumPy or Pandas functions.

e. Statistical Aggregation: Based on the results from steps b, c, and d, maximum, minimum, and average values are computed for each feature. These aggregated values offer a consolidated view of a feature's importance and help in objective decision-making during the feature selection process.

f. Feature Selection and Normalization: A final set of features is selected based on the above evaluations. This step eliminates irrelevant and redundant features, optimizing model performance and reducing computational complexity. After selection, the dataset is normalized (e.g., Min-Max scaling or Z-score standardization) to bring all features to a uniform scale. Scikit-learn's preprocessing module is utilized for this step.

g. Data Splitting: The normalized dataset is split into training and testing sets, typically in an 80:20 ratio. The training data is used to build the model, while the test data is held out for evaluating the model's generalization capability.

h. Model Training with ML Algorithms: Three machine learning models are trained using the selected features:

- **Logistic Regression (LR)** – A statistical model for binary classification.
- **Support Vector Machine (SVM)** – An effective classifier for high-dimensional data.
- **Multi-Layer Perceptron (MLP) –** A feedforward artificial neural network with one or more hidden layers.

These models are trained using Scikit-learn and Keras libraries. Hyperparameter tuning and cross-validation techniques are applied to optimize performance.

i. Model evaluation and prediction: The trained models are evaluated using the test dataset. Standard evaluation metrics such as accuracy, confusion matrix, precision, recall, and F1-score are calculated. These metrics assess the performance of each model in predicting CKD and guide the selection of the best-performing algorithm.

2) Feature Reduction: The dataset includes 24 independent variables. Models learn faster and more effectively when features are coherent and relevant, whereas excessive features increase learning and computational complexity. Thus, reducing fewer contributing features is common in ML. This study employs a novel feature reduction method using heatmaps, information gain, and standard deviation.

a) Heatmap Analysis: Heatmap analysis was used to visualize feature correlations, reveal patterns or redundancies, and retain only features with unique contributions. Each cell in the heatmap represents a data point, with values calculated using Eq. 1 (Fig. 2).

$$\mathrm{Heatmap(i,j)} = \frac{Data(i,j) - Min(Data)}{Max(Data) - Min(Data)} \quad \ldots (1)$$

The data value at point (i,j) in the matrix is represented by Data (i,j) in this case. The equation presents a value between 0 and 1 for each of the two features.

b) Information Gain: To assess each characteristic in the dataset's relevance to the target variable, information gain, a metric originating in the discipline of information theory was applied. According to (1), features with greater information values were deemed more informative for the prediction task.

In machine learning and decision tree methods, the information gain formula is frequently utilized to assess how well a feature classifies a dataset. The formula of information gain is typically associated with the entropy measure. Given a dataset D and a feature A, the information gain (IG) is calculated by Eq. 2.

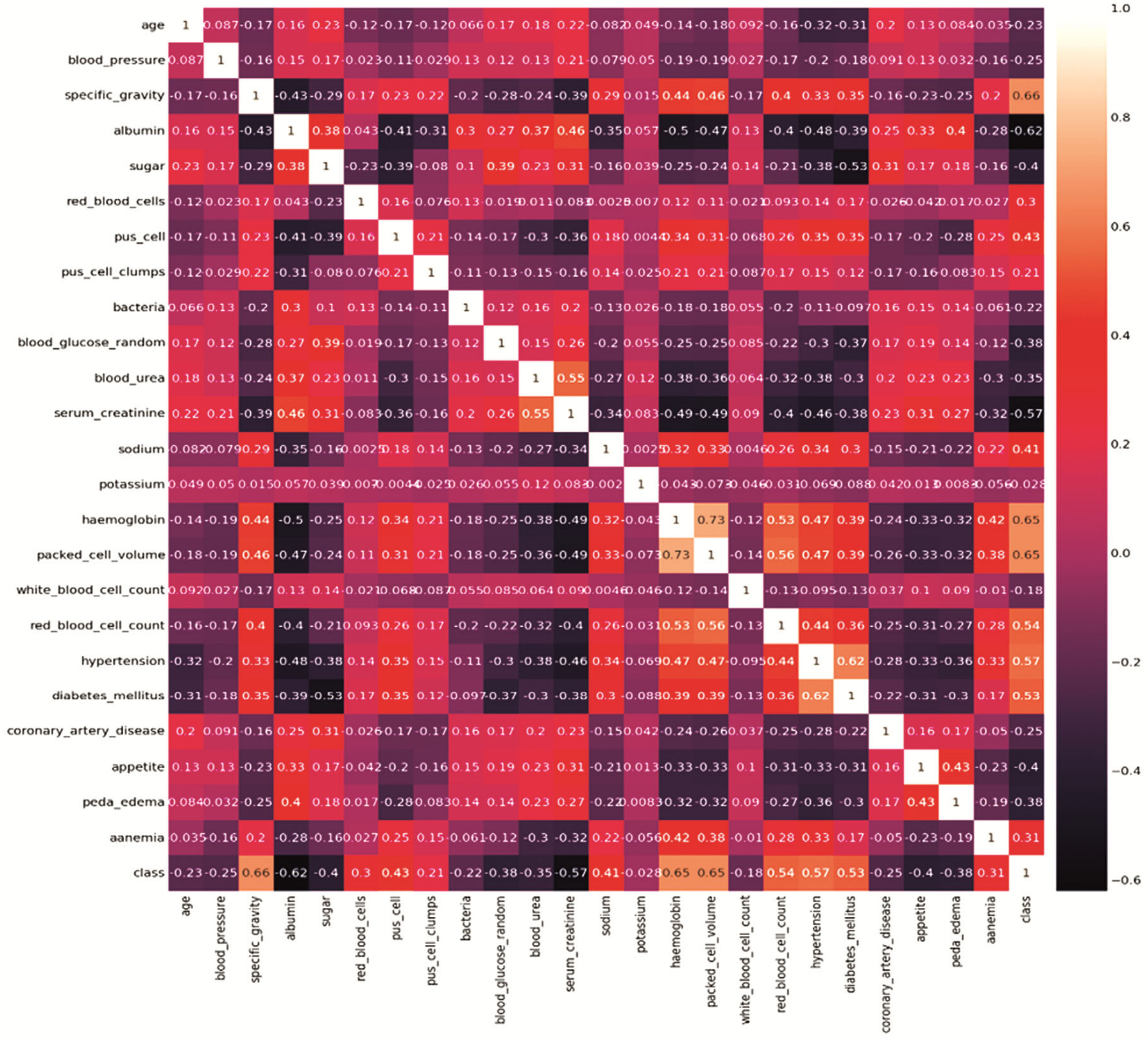

Fig. 2 — Heatmap

$$IG(D,A) = H(D) - H(D|A) \qquad \dots (2)$$

where, the complete dataset is *D*, entropy is represented by *H*(*D*) and the conditional entropy of *D* given the feature *A* is represented by *H*(*D*|*A*). The entropy *H* is calculated by Eq. 3:

$$\mathrm{H(D)} = \sum_{\mathrm{i=1}}^{\mathrm{c}} \mathrm{p_i} \cdot \log_2(\mathrm{p_i}) \qquad \dots (3)$$

where, *c* is the number of classes in the dataset, and p_i is the proportion of instances in the dataset *D* that correspond to class *i*.

The conditional entropy *H*(*D*|*A*) is calculated by Eq. 4 as: $\mathrm{D_j}$

$$H(D|A) = \sum_{j=1}^{k} \frac{|\mathrm{Di}|}{|D|} \cdot H(Dj) \qquad \dots (4)$$

where, *k* is the number of unique values that feature *A* can take, D_j is the subset of *D* where feature *A* has the *j*-th value, and $|D_j|$ is the size of D_j.

In practical terms, a higher information gain suggests that the feature is more valuable in decreasing the uncertainty about the class labels in the dataset. To make the calculation easier with information gain, *IG*(*D* < *A*) can be addressed in the following as *IG*(*i*), where $1 \leq i \leq k$ and k is the number of features.

c) Standard Deviation: Standard deviation was used to measure data dispersion within each feature. Features with low variability were considered for removal, as they likely offer less discriminatory value. If a set of *n* data points $x_1, x_2, \ldots, x_n$ of a feature *i* and its mean *x* are given, the standard deviation (*σ*(*i*) of data of feature *i* is measured by Eq. 5 as:

$$\sigma(i) = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}} \qquad \dots (5)$$

d) Computing the values for feature selection parameters: This process selects the most relevant features for accurate CKD detection, known as feature reduction. The heatmap (Eq. 1) provides a 2D correlation matrix, while information gain and standard deviation yield a single value per feature. To combine these measures, the dataset was first reformatted (Table 1).

Making relation, i.e., integrating their results in a relation, some of them can be resized to set a homogeneous dimension. For this, the row values of Heatmap are averaged and then the averaged values are stored to *HM*. The HM is generated using Eq. 6, which is a one-dimensional array of data structures.

$$HM(i) = \frac{\sum_{j=1}^{k} Heatmap(i,j)}{k} \quad \dots (6)$$

where, k is the number of features.

At that stage of computation, it can be seen that *HM*, *IG,* and σ are now analogous by their dimension and numerical values. The average, minimum, and maximum values of *HM*, *IG,* and σ for each feature might help us to gain insights into their distributional characteristics, aiding in the identification of potential outliers and anomalies.[22] To justify that argument, the feature-wise maximum, minimum, and average values were measured from HM, IG, and σ by Eqs 7, 8, and 9 in F_{Max}, F_{Min}, and F_{Avg}, respectively.

$$F_{Max}(i) = \max\,(HM(i), IG(i), \sigma(i)) \quad \dots (7)$$

$$F_{Min}(i) = \min\,(HM(i), IG(i), \sigma(i)) \quad \dots (8)$$

$$F_{avg}(i) = avg(HM(i), IG(i), \sigma(i)) \quad \dots (9)$$

where, the function $max(a,b,c)$, $min(a,b,c)$, and $avg(a,b,c)$ return the maximum, minimum, and average of a,b and c, respectively.

e) Feature Selection: Next, F_{Min}, F_{Min}, and F_{Avg}, were dealt separately. F_{Max} contains a single value for each feature. To avoid the fewer contributing features from applying in the final model, the lowest valued feature in F_{Max} were dropped. The remaining features are then called selected features and have been used in current machine learning models. The feature-dropping procedure is repeated from the selected features until the performance of the model does not decrease. In that repetition process, after every drop of features, the selected features are set as a queue of entering the intended machine learning models. However, a normalization of feature values is performed before using in machine learning model. That normalization

Table 1 — Combination of Heatmap, Information gain, and Standard deviation

Feature	Feature's code	Heatmap	Information gain	Standard deviation	Feature wise maximum, minimum and average of the parameters		
					Minimum	Maximum	Average
Age	V1	0.767	0.241	0.617	0.2409	0.7672	0.5549
Blood pressure	V2	0.720	0.221	0.599	0.2210	0.7195	0.5149
Specific gravity	V3	0.065	0.550	0.000	0.0000	0.5498	0.2185
Albumin	V4	1.000	0.478	0.419	0.4187	1.0000	0.6307
Sugar	V5	0.886	0.171	0.402	0.1712	0.8863	0.4964
Red blood cells	V6	0.501	0.389	0.368	0.3679	0.5009	0.4600
Pus cell	V7	0.703	0.216	0.360	0.2163	0.7028	0.4440
Pus cell clumps	V8	0.845	0.089	0.314	0.0891	0.8450	0.4348
Bacteria	V9	0.783	0.052	0.294	0.0521	0.7833	0.3938
Blood glucose random	V10	0.850	0.401	0.731	0.4011	0.8496	0.6667
Blood urea	V11	0.882	0.346	0.699	0.3456	0.8822	0.6588
Serum creatinine	V12	0.941	0.564	0.532	0.5324	0.9409	0.6744
Sodium	V13	0.156	0.395	0.570	0.1560	0.5701	0.3673
Potassium	V14	0.597	0.351	0.480	0.3506	0.5967	0.4867
Haemoglobin	V15	0.005	0.709	0.477	0.0046	0.7087	0.3976
Packed cell volume	V16	0.000	0.647	0.561	0.0000	0.6465	0.4025
White blood cell count	V17	0.622	0.316	1.000	0.3160	1.0000	0.6403
Red blood cell count	V18	0.051	0.629	0.387	0.0506	0.6286	0.3492
Hypertension	V19	0.995	0.342	0.346	0.3416	0.9945	0.5625
Coronary artery disease	V21	0.808	0.067	0.305	0.0675	0.8077	0.4109
Appetite	V22	0.870	0.164	0.332	0.1640	0.8701	0.4627
Pedal edema	V23	0.843	0.150	0.330	0.1504	0.8428	0.4460
Anaemia	V24	0.746	0.116	0.322	0.1159	0.7460	0.3894

technique is described in the following sub-section. The same methodologies are exercised for F_{Min}, and F_{Avg} as well, and every time, the statics of performance measuring parameters are noted down.

3) Normalization and Statistical Measures: Following the feature selection process, the values of the selected features were normalized to a desired extent to ensure consistent scales for subsequent analyses. Normalization aims to ensure that each feature makes an equal contribution to the analysis and stops some features from taking center stage because of their size. When working with algorithms that are sensitive to the scale of input data, such as distance-based approaches and gradient-based optimization algorithms, normalization is very crucial.

There are several types of normalization methods, e.g., MinMax, Z-Score, t-score, frequency distribution, or probability distribution-based normalization. However, Min-Max normalization is used for this study.

Min-Max Scaling (Min-Max Normalization): A normalization method called Min-Max scaling sets the data's scale of each feature to a predetermined range, usually between 0 and 1. The following Eq. 10 is the formula for Min-Max Scaling. That formula is applied to the data of each feature, individually.

$$X_{normalized} = \frac{X - \min(X)}{\max(x) - \min(X)} \quad \ldots (10)$$

In that case, the minimum and maximum values in a working feature are computed by min(X) and max(X), respectively, while X represents the set of data points of the same feature.

4) Machine Learning Models: To evaluate the impact of feature selection on expected performance, three popular machine learning algorithms were used: A methodical methodology was used to iteratively remove each of the three algorithms—Logistic Regression (MLP), Support Vector Machine, and Multi-Layer Perceptron based on their normalized statistical measures and evaluate the change in the model accuracy that resulted.

Logistic Regression: To determine the ideal collection of features maximizing prediction accuracy, logistic regression, a popular linear model for binary classification problems, was trained and assessed using several feature subsets.[23] In logistic regression, the logistic function—also called the sigmoid function—is used to determine the probability of a binary outcome. The expression for the Logistic Regression formula, i.e. Eq. 11, is:

$$P(Y=1) = \frac{1}{1+e^{-(b_0+b_1X_1+b_2X_2+\cdots+b_nX_n)}} \quad \ldots (11)$$

Support Vector Machine: Using a similar iterative feature removal procedure, the dataset was analyzed using SVM, a potent learning model suitable for both linear and non-linear data. The objective was to identify the subset of features that would result in the highest predictive accuracy.[24] Alternatively, it can be stated that the SVM algorithm is employed for both classification and regression tasks. The fundamental concept of classification is represented by the hyperplane, which serves as the most effective means of dividing the data into distinct classes. For a two-class SVM, the equation of the hyperplane is expressed as follows:

$$f(x) = \sin\left(\textstyle\sum_{i=1}^{n} w_i x_i + b\right) \quad \ldots (12)$$

The SVM's training goal is to determine the optimal values for w and b that maximize the margin between the two classes while reducing the classification error.

For, a linearly separable dataset the decision function can also be expressed in terms of the weight vector and feature vector x inner product w:

$$f(x) = \sin(\langle w, x\rangle + b) \quad \ldots (13)$$

(w,x) here, indicates the dot product of the functions w and x.

In practice, methods such as the kernel trick, which entails projecting the input features into a higher dimensional space, are frequently used to expand the SVM formulation for nonlinearly separable datasets. When using kernelized SVM, the function becomes:

$$f(x) = \sin(\textstyle\sum_{i=1}^{n} a_i y_i K(x, x_{i)} + b \quad \ldots (14)$$

where, the class label is y_i, the Lagrange multipliers are α_i, the dot product is $\langle w,x\rangle$, and the kernel function is $K(x, x_i)$.

Multi-Layer Perceptron: A neural network architecture called MLP, was employed to capture complex relationships within the dataset. Similar to other models, feature selection iterations were performed to optimize the predictive performance of the MLP classifier.

The equation for a multi-layer perceptron, a kind of artificial neural network, has the following mathematical definition. Let us consider a simple single-layer feed-forward network. The input layer has m neurons, the hidden layer contains h neurons, and the output layer contains k neurons. The weight matrix that links the input layer and hidden layer is $W^{(1)}$; the weight matrix that links the hidden layer and output layer is $W^{(2)}$. Moreover, $b^{(1)}$ and $b^{(2)}$ represent the bias vectors for the output and hidden layers, respectively.

Let x stand for the input vector, $a^{(1)}$ for the output of the hidden layer after the activation function, and $a^{(2)}$ for the final output of the network. One way to write the equations for a basic MLP is as follows:

$$z^{(1)} = W^{(1)} + b^{(1)} \quad \dots (15)$$

$$A^{(1)} = \sigma(z^{(1)})(W\square ere\sigma ist\square eactivationfunction) \quad \dots (16)$$

$$z^{(2)} = W^{(2)}a^{(1)} + b^{(2)} \quad \dots (17)$$

$$a^{(2)} = \sigma(z^{(2)}) \quad \dots (18)$$

In matrix form:

$$z^{(1)} = W^{(1)}x + b^{(1)} \quad \dots (19)$$

$$a^{(1)} = \sigma(z^{(1)}) \quad \dots (20)$$

$$z^{(2)} = W^{(2)}a^{(1)} + b^{(2)} \quad \dots (21)$$

$$a^{(2)} = \sigma\left(z^{(2)}\right) \quad \dots (22)$$

These equations describe the forward pass of the neural network, where the input x is propagated through the network to produce the final output $a^{(2)}$. Numerous alternatives for the activation function σ and the network's distinctive design, which includes the number of hidden layers and neurons in each layer, may be taken into consideration depending on the nature of the problem.

Experimental Evaluation

First of all, the average values of each feature was calculated the average value of each feature by using these three measurements

$$A[i] = \frac{H[i]+S[i]+IG[i]}{3} \quad \dots (23)$$

This formula averages the scores obtained from each methodology, ensuring a balanced consideration of the insights derived from heatmap visualization, standard deviation analysis, and information gain assessment.

Secondly, the maximum values for each feature were determined using these three measurements.

Finally, the minimum values were identified using all these three measurements.

Impact of Feature Removal on Model Accuracy

Scenario 1: Using averaged values Logistic Regression examined the impact of removing features on model accuracy. Features 'age,' 'age and blood pressure,' and 'age, blood pressure, and specific gravity' were sequentially removed, retraining and evaluating the models for each step, as shown in Table 2. Remarkably, all three ML models achieved 100% accuracy after these removals (Fig. 3). This highlights possible redundancy or collinearity among these features in CKD prediction. The presence of null values is shown in Fig. 4, while precision, recall, and F1-score for average values are presented in Fig. 5.

Scenario 2: Using maximum values Upon removing the features 'age,' 'age and blood pressure,' and 'age, blood pressure, and specific gravity' for maximum values (Fig. 6), the Logistic Regression model achieved an impressive accuracy of 97.50%, as in Table 3. This observation highlights the effectiveness of feature removal in enhancing model accuracy, especially when addressing potentially redundant or co-linear features.

Scenario 3: Using minimum values removing only the 'age' feature improved accuracy to 98.75% (Table 4, Fig. 7). This suggests that 'age' may introduce noise in this context. Building on preprocessing and feature importance insights, MLP, SVM, and Logistic Regression models were trained on 80% of the CKD dataset, with 20% used for testing. Logistic Regression with average feature values achieved the best results. A combined confusion matrix of all three scenarios visualizes model performance (Fig. 8). Comparative analyses with prior studies are shown in Tables 4 and 5.

Comparative Analysis for Logistic Regression

Validation of the Proposed Work with Recent Studies

1) Recent studies highlight feature selection's role in improving accuracy. Study 1(17) used an ensemble with Random Forest, achieving 98% accuracy. Study 3(15) applied multiple selection techniques, reaching 100% accuracy with Random Forest and AdaBoost. The current study similarly employed information gain, heatmap analysis, and standard deviation, contributing to high model accuracy.

Table 2 — List of reduced features by the recursive reduction process

No. of reduced features	BY maximum	BY minimum	BY average
1	V6	V3	V23
2	V6, V3	V3, V16	V23, V24
3	V6, V3, V13	V3, V16, V15	V23, V24, V5
4	V6, V3, V13, V14	V3, V16, V15, V18	V23, V24, V5, V22
5	V6, V3, V13, V14, V18	V3, V16, V15, V18, V9	V23, V24, V5, V22,V20
6	V6, V3, V13, V14, V18,V16	V3, V16, V15, V18, V9,V21	V23, V24, V5, V22,V20,V21
7	V6, V3, V13, V14, V18,V16, V7	V3, V16, V15, V18, V9,V21, V8	V23, V24, V5, V22,V20,V21,V3
8	V6, V3, V13, V14, V18,V16, V7, V15	V3, V16, V15, V18, V9,V21, V8, V23	V23, V24, V5, V22,V20,V21,V3,V19
9	V6, V3, V13, V14, V18,V16, V7, V15, V2	V3, V16, V15, V18, V9,V21, V8, V23, V23	V23, V24, V5, V22,V20,V21,V3,V19,V8
10	V6, V3, V13, V14, V18,V16, V7, V15, V2, V24	V3, V16, V15, V18, V9, V21, V8, V23, V23, V13	V23, V24, V5, V22,V20,V21,V3,V19,V8,V9
11	V6, V3, V13, V14, V18,V16, V7, V15, V2, V24, V1	V3, V16, V15, V18, V9,V21, V8, V23, V23, V13, V22	V23, V24, V5, V22,V20,V21,V3,V19,V8,V9, V2
12	V6, V3, V13, V14, V18,V16, V7, V15, V2, V24, V1, V9	V3, V16, V15, V18, V9,V21, V8, V23, V23, V13, V22, V5	V23, V24, V5, V22,V20,V21,V3,V19,V8,V9, V2,V4
13	V6, V3, V13, V14, V18,V16, V7, V15, V2, V24, V1, V9,V21	V3, V16, V15, V18, V9,V21, V8, V23, V23, V13, V22, V5, V7	V23, V24, V5, V22,V20,V21,V3,V19,V8,V9, V2,V4,V7
14	V6, V3, V13, V14, V18,V16, V7, V15, V2, V24,V1, V9, V21, V23	V3, V16, V15, V18, V9,V21, V8, V23, V23, V13,V22, V5, V7, V2	V23, V24, V5, V22,V20,V21,V3,V19,V8,V9, V2,V4,V7,V1

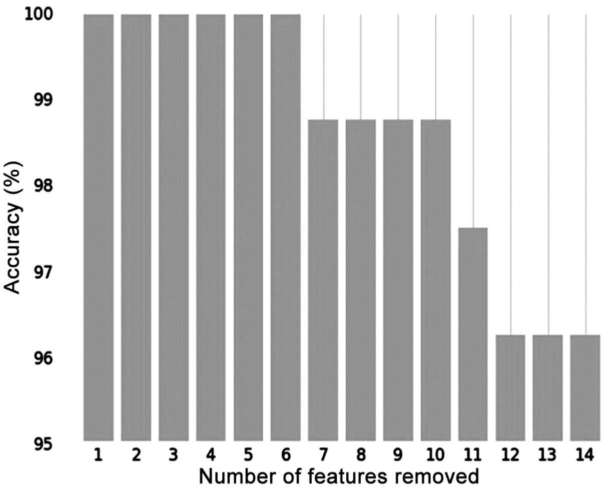

Fig. 3 — Accuracies of average values

2) While prior work showed feature selection's impact, the current study advances this by introducing new metrics for feature importance. Accuracy of 100% (Fig. 7) is achieved by removing nine features based on these metrics, validating feature selection's importance and offering a novel approach to improve model robustness.
3) Study 2(16) reported improved accuracy after feature selection, consistent with findings of the current study. Both emphasize selecting critical features to boost performance; Accuracy of 100% of this study further corroborates this.

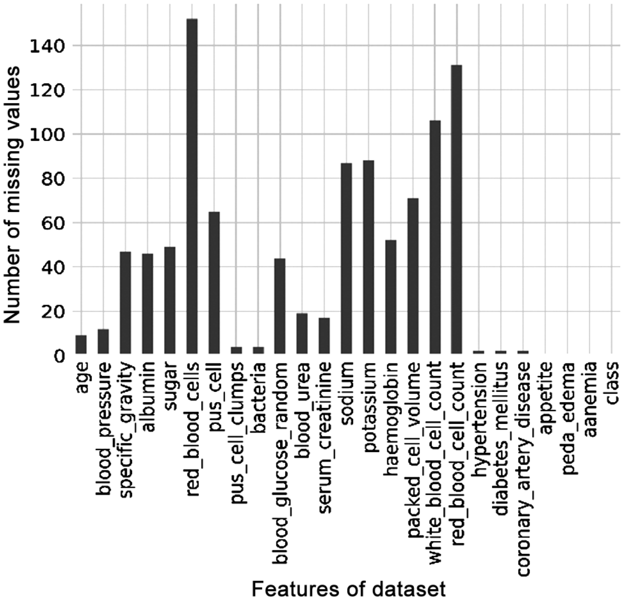

Fig. 4 — Presence of null values

Novelty and Significance of the Proposed Method

The study introduces three novel metrics—heatmap analysis, information gain, and standard deviation—to evaluate feature importance in CKD prediction. Unlike conventional single-metric approaches, a comprehensive framework was developed to assess their combined impact on model performance, enhancing both accuracy and interpretability. A comparative analysis using Logistic Regression, Support Vector Machine, and Multi-Layer Perceptron showed that the average metric achieved the highest

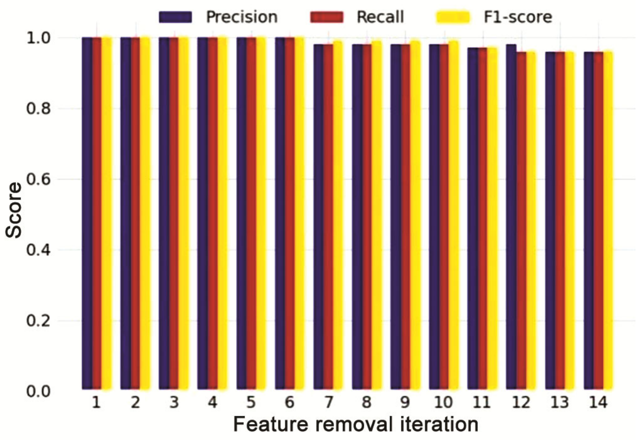

Fig. 5 — Precision, Recall and F1-score

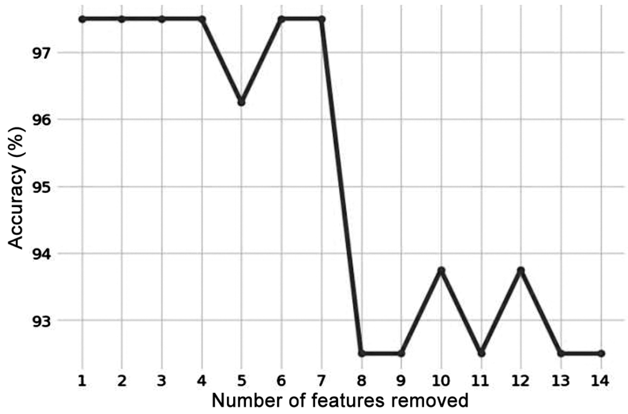

Fig. 6 — Accuracies of maximum values

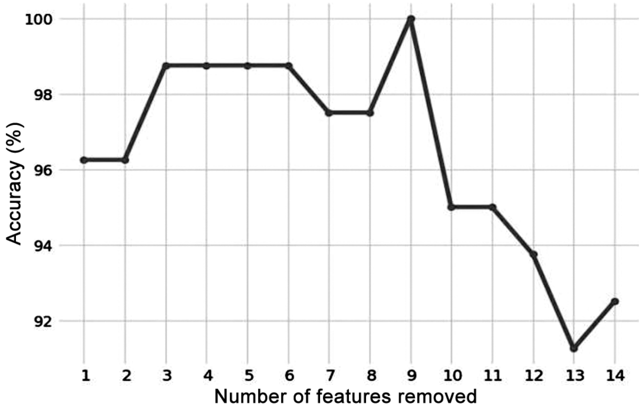

Fig. 7 — Accuracies of minimum values

Table 3 — Results of performance measuring parameters for recursive reduction polices

No. of reduced features	(Maximum, Minimum, Average)		
	F1 Score	Recall	Accuracy
1	100, 96, 100	100, 95, 100	97.50, 96.25, 100
2	100, 96, 100	100, 95, 100	97.50, 96.25, 100
3	100, 99, 100	100, 99, 100	97.50, 98.75, 100
4	100, 99, 100	100, 99, 100	97.50, 98.75, 100
5	97, 99, 100	98, 99, 100	96.25, 98.75, 100
6	97, 99, 100	94, 99,100	97.50, 98.75, 100
7	97, 98, 99	94, 98,100	97.50, 97.50, 98.75
8	99, 98, 99	100, 98,100	92.50, 97.50, 98.75
9	93, 100, 99	97, 100, 100	92.50, 100, 98.75
10	93, 95, 99	100, 96, 100	93.75, 95, 98.75
11	93, 95, 98	100, 96, 98	92.50, 95, 97.50
12	94, 94, 97	100, 94, 96	93.75, 93.75, 96.25
13	93, 90, 97	100, 92, 96	92.50, 91.25, 96.25
14	92, 91, 97	97, 93, 98	92.50, 92.50, 96.25

accuracy, while the minimum metric reduced complexity. Overall, the proposed method improved early CKD detection and supported personalized, cost-effective healthcare.

Future Scope

This study opens several directions for future research and practical implementation in CKD prediction. Machine-learning models should be refined for real-world healthcare settings, ensuring smooth integration with Electronic Health Records (EHR) and usability for clinicians with minimal training. The proposed models and metrics also require validation on larger and more diverse datasets to confirm their consistency across different populations and clinical environments. In addition to MLP, SVM, and LR, future studies should explore algorithms such as Random Forests, Gradient Boosting, and other ensemble techniques to further improve accuracy and robustness. Incorporating longitudinal data could provide deeper insights into CKD progression, enabling early detection and timely interventions. Moreover, developing real-time predictive tools based on the proposed models could support instant CKD risk assessments within clinical workflows. Finally, enhancing the transparency and interpretability of complex models, particularly MLP,

Table 4 — Comparison with different studies

Concepts and techniques	Classifier	Accuracy %	Precision %	Recall %	F1 Score %
Information gain[1]	CS adaboosting	99.8	100	98	99
Recursive feature elimination[12]	xgboost	98.30	98	98	98
Bagging ensemble technique[19]	Decision tree	97.23	98	99	99
Information gain, Heatmap and standard deviation[This Study]	Logistic regression (With minimum values after removing 9 features)	100	100	100	100

Table 5 — Comparative analysis of recent studies and the current work on CKD prediction

Criteria	Study 1 (Ensemble model)	Study 2 (Multiple ML methods)	Study 3 (ML-CKDP model)	This work
Study focus/Objective	Predict CKD using an Ensemble model combining top-performing deep learning models.	Early identification of CKD using multiple ML methods (KNN, RF, DT, GB, XG-Boost).	Develop a web-based CKD prediction model with comprehensive preprocessing and classifier evaluation.	Introduce new metrics for feature importance evaluation and predict CKD using MLP, SVM, and LR models.
Feature selection	Applied Random Forest feature selection to reduce the number of features.	Exhaustive feature selection using Wrapper methodology.	Used multiple feature selection techniques (Correlation, ChiSquare, RFE, Lasso, Ridge).	Removed 9 features based on standard deviation and information gain, achieving 100% accuracy.
Feature reduction	Specific methods for feature reduction are not explicitly detailed.	KNN improved accuracy from 77% to 83% after feature reduction.	Comprehensive feature selection with no specific feature reduction method mentioned.	Reduced features such as specific gravity, packed cell volume, hemoglobin, etc., achieving 100% accuracy after reduction.
Accuracy achieved	98% (6 months dataset), 97% (12 months dataset) using Ensemble Model.	KNN improved from 77% to 83%. No other models are specified for accuracy after feature selection.	RF and AdaBoost achieved 100% accuracy.	100% for average values (after removing the first feature), 100% for minimum values (after removing nine features), and 96.75% for maximum values (after removing one feature).
Practical application	Focus on predicting CKD before it occurs, with the potential to lower healthcare costs.	Aims to encourage early diagnosis of CKD for adopting a healthy lifestyle.	Web-based application for real-time CKD prediction.	Potential to enhance early CKD detection and management, reducing healthcare costs and improving patient outcomes.
Limitations	Did not explicitly mention limitations regarding feature selection or model complexity.	Focus on accuracy improvement; less emphasis on real-world application.	High computational requirements; performance across different datasets may vary.	Computational resource demands, potential for overfitting in smaller datasets.

will be crucial for building clinical trust and supporting informed decision-making.

Real-Life Applications

The proposed CKD prediction model has several practical applications that can significantly impact healthcare delivery and patient outcomes.

1) Early diagnosis and intervention, enable healthcare providers to detect CKD at an earlier stage and implement preventive measures, such as lifestyle modifications or medications, to slow disease progression.
2) Personalized treatment plans tailored to an individual's specific risk profile. This approach enhances the effectiveness of treatment, reduces the need for invasive procedures, and ultimately improves patient outcomes.
3) Remote monitoring and telemedicine, the model can be integrated into telemedicine platforms, enabling the remote monitoring of patients and early detection of CKD, especially in rural or underserved areas. This application increases accessibility to critical healthcare services and ensures timely interventions.
4) Electronic Health Records (EHR) systems, where healthcare providers receive real-time alerts for patients at high risk for CKD, facilitating proactive management and improving clinical decision-making.

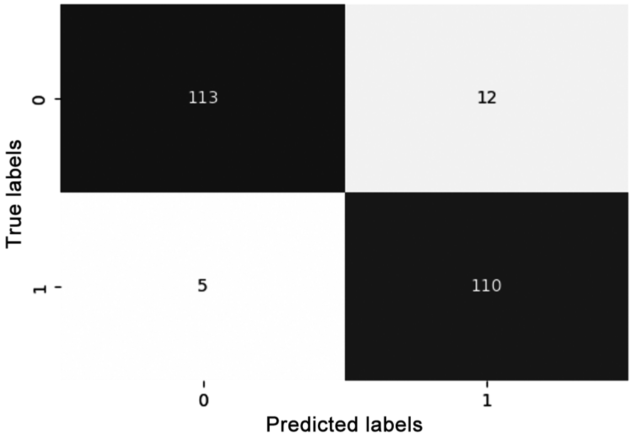

Fig. 8 — Merged confusion matrix of MLP, LR and SVM

5) Insurance and risk assessment, the model can be used by insurance companies to assess CKD risk in policyholders, leading to better risk management, more accurate premium calculations, and encouraging preventive health measures among individuals.

Limitation of Proposed Work

A key limitation of this study is the use of a single dataset for developing, training, and evaluating the CKD prediction models. This may restrict the generalizability of the results, as the model's performance could vary when applied to other populations or clinical settings with different demographic, genetic, or environmental profiles.

Conclusions

This study significantly advances CKD prediction by introducing three new metrics for evaluating feature importance: information gain, heatmap analysis, and standard deviation. These metrics offer valuable insights into the relative significance of different features. Using them, three machine learning models—Multi-Layer Perceptron (MLP), Support Vector Machine (SVM), and Logistic Regression (LR)—achieved excellent results: 100% accuracy for average values (after removing the first feature), 100% for minimum values (after removing nine features), and 96.75% for maximum values (after removing one feature). These findings demonstrate the robustness and reliability of the proposed approach. However, the study is limited by high computational requirements and possible overfitting on smaller datasets. Future work will focus on optimizing these models for clinical applications and testing them on larger datasets. Overall, the study enhances early CKD detection and management, supporting better patient outcomes and reducing healthcare costs.

Acknowledgement

This research work was supported by the ICT Division, Ministry of Posts, Telecommunications and Information Technology, Bangladesh.

References

1 Alloghani M, Al-Jumeily A, Hassan H, Al-Hussein L & Al-Bayatti A, Performance-based prediction of chronic kidney disease using machine learning for high-risk cardiovascular disease patients, *Stud Comput Intell*, **855** (2020) 187–206, DOI: 10.1007/978-3-030-28553-1_9.

2 Almansour N A, Syed H F, Khayat N R, Juri R J, Altheeb A S, Alhiyafi J A, Alrashed S & Olatunji S O, Neural network and support vector machine for the prediction of chronic kidney disease: A comparative study, *Comput Biol Med*, (2019) 101–111, DOI: 10.1016/j.compbiomed.2019.04.017.

3 Ampavathi A & Saradhi T V, Multi disease-prediction framework using hybrid deep learning: an optimal prediction model,*Comput Methods Biomech Biomed Eng*, **24(10)** (2020) 1146–1168, DOI: 10.1080/10255842.2020.1869726.

4 Bai Q, Su C, Tang W & Li Y, Machine learning to predict end stage kidney disease in chronic kidney disease, *Sci Rep*, **12(1)** (2022) 8377, DOI: 10.1038/s41598-022-12316-z.

5 Battineni G, Naik N, Amenta F & Cavallucci G, Applications of machine learning predictive models in the chronic disease diagnosis, *J Pers Med*, **10(2)** (2020) 21, DOI: 10.3390/jpm10020021.

6 Ebiaredoh-Mienye S A, Swart T G, Esenogho E & Mienye I D, A machine learning method with filter-based feature selection for improved prediction of chronic kidney disease, *Bioeng*, **9(8)** (2020) 350, DOI: 10.3390/bioengineering 9080350.

7 Fatema K, Khatun A A, Brisrty S A, Islam M A, Kabir F N A, Biswas Nayan S, Afrin A, Sathi T R & Halim M A, Analytical studies on chronic kidney disease and associated lifestyle risk factors among the patients of different hospital in Dhaka city, Bangladesh, *Bangladesh J Food Nutr*, **8** (2025) 1–15, DOI: 10.9734/ajrn/2025/v8i187.

8 Gudeti B, Mishra S, Malik S, Fernandez T F, Tyagi A K & Kumari S, A novel approach to predict chronic kidney disease using machine learning algorithms, *Int Conf Electron Commun Aerosp Technol (ICECA)*, **4** (2020) 1630–1635, DOI: 10.1109/ICECA49313.2020.9297392.

9 Halder R K, Hossain M S, Saha S, Rahim A, Asifuzzaman M & Alam M, ML-CKDP: Machine learning-based chronic kidney disease prediction with smart web application, *J Pathol Inform*, **15(1)** (2024) 100371, DOI: 10.1016/ j.jpi.2024.100371.

10 Hema K & Prasanna R, Analyze the impact of feature selection techniques in the early prediction of CKD, *Int J Coan Comput Eng*, **5** (2024) 66–77, DOI: 10.1016/ j.ijcce.2023.12.002.

11 Islam M A, Ahamed S H, Malik K S, Saba T S & Hasan S, Risk factor prediction of chronic kidney disease based on

machine learning algorithms, *Int Conf Intell Sustain Syst (ICISS),* **3rd** (2020) 952–957, DOI: 10.1109/ ICISS49785.2020.9316044.

12 Islam M M & Islam H M, Chronic kidney disease prediction based on machine learning algorithms, *J Pathol Inform*, **14(100189)** (2022) DOI: 10.1016/j.jpi.2023.100189.

13 Jena L, Patra B, Nayak S, Mishra S & Tripathy S, Risk prediction of kidney disease using machine learning strategies, *Springer*, **153** (2020) 485–494, DOI: 10.1007/978-981-15-6202-0_50.

14 Nishat M M, Faisal F, Dip R R, Sarker M N, Ahsan R, Shikder F, Asif M A & Hoque M A, A comprehensive analysis on detecting chronic kidney disease by employing machine learning algorithms, *EAI Endors Trans Pervas Health Technol*, **7** (2021) 1–12, DOI: 10.4108/eai.13-8-2021.170671.

15 Pal S, Chronic kidney disease prediction using machine learning techniques, *Biomed Mater Devices*, **1(1)** (2023) 534–540, DOI: 10.1007/s44174-022-00027-y.

16 Polat H, Danaei Mehr H & Çelebi A, Diagnosis of chronic kidney disease based on support vector machine by feature selection methods, *J Med Syst,* (2017) 111, DOI: 10.1007/s10916-017-0711-3.

17 Sharma A & Rani R, Drug sensitivity prediction framework using ensemble and multi-task learning, *Int J Mach Learn Cyber*, **11** (2020) 1231–1240, DOI: 10.1007/s13042-019-01034-0.

18 Radovic N, Vitale P, Emin M V & Antunovic T, Machine learning approach in mortality rate prediction for hemodialysis patients, *Comput Methods Biomech Biomed Eng*, **25(1)** (2022) 111–112, DOI: 10.1080/ 10255842. 2021.1937611.

19 Rashed-Al-Mahfuz M, Abdur Rahman M, Ahmed S, Al Mamun A, Quinn J M W & Moni M A, Clinically applicable machine learning approaches to identify attributes of chronic kidney disease (CKD) for use in low-cost diagnostic screening, *IEEE J Trans Eng Health Med*, **9** (2021) 1–11.

20 Sahoo P & Sahoo P, Intelligent and cloud computing, *Int J Eng Res Technol (ICICC 2019 Proc)*, **8** (2019) 608–612, DOI: 10.1007/978-981-15-5679-1_55.

21 Saif D, Sarhan A M & Elshennawy N M, Early prediction of chronic kidney disease based on ensemble of deep learning models and optimizers, *J Electr Syst Inf Technol*, **11(1)** (2024) 17, DOI: 10.1186/s43067-024-00142-4.

22 Sanmarchi F, Cortesi P A, Golinelli D, Di Gaetano G, Esposito D, Held-Baloncieri T & Capodici A, Predict, diagnose, and treat chronic kidney disease with machine learning: a systematic literature review, *J Nephrol*, **36(4)** (2023) 1101–1117, DOI: 10.1007/s40620-023-01573-4.

23 Wang X, Yang W, Zhang Y, Sun S, Yu L & Xu S, Analysis and prediction of gestational diabetes mellitus by the ensemble learning method, *Int J Comput Intell Syst*, **15** (2022) 72, DOI: 10.1007/s44196-022-00110-8.

24 Xiao J, Ding R, Xu X, Guan H, Feng X, Sun T, Zhu S & Ye Z,Comparison and development of machine learning tools in the prediction of chronic kidney disease progression, *J Transl Med*, **17** (2019) 119, DOI: 10.1186/s12967-019-1860-0.