International Journal of Statistical Sciences *ISSN 1683-5603*
Vol. 25(2), November, 2025, pp 1-32
DOI: https://doi.org/10.3329/ijss.v25i2.85732
© 2025 Dept. of Statistics, Univ. of Rajshahi, Bangladesh

Identification of Predisposing Risk Factors for Chronic Kidney Disease and Optimizing Disease Prediction Using a Stacking Machine Learning Algorithm

Md. Razu Ahmed[1], Md. Abdur Rakib[1], Abu Bakar Shiddik[1], and Md. Shamim Reza[2 *]

[1]Deep Statistical Learning & Research Lab, Department of Statistics, Pabna University of Science & Technology, Pabna-6600, Bangladesh

[2]Department of Statistics, Pabna University of Science & Technology, Pabna-6600, Bangladesh

[*]Correspondence should be addressed to Md. Shamim Reza
(Email: shamim.reza@pust.ac.bd)

[Received May 1, 2025; Accepted October 26, 2025]

Abstract

Chronic kidney disease (CKD) remains a major global health concern, often progressing silently until advanced stages. Early detection is therefore critical to improving outcomes. This study focuses on identifying the most important predisposing risk factors for CKD and optimizing prediction performance using a stacking machine learning ensemble. Two datasets were analyzed: the UCI CKD dataset and a synthetically simulated CKD dataset designed to mirror real-world variability. A leakage-safe preprocessing pipeline was implemented, including median and mode imputation for missing values, Z-score capping for outliers, min–max normalization, and class balancing through the Synthetic Minority Over-sampling Technique (SMOTE). Feature selection was performed using six complementary approaches: Logistic Regression (LR), Recursive Feature Elimination (RFE), Random Forest (RF), Mutual Information (MI), Chi-Square (χ^2), and Principal Component Analysis (PCA) with a majority-vote strategy used to identify features consistently recognized as predictive. For the UCI CKD dataset, the common risk factors were serum creatinine, hemoglobin, packed cell volume, red blood cell count, specific gravity, albumin, sugar, hypertension, diabetes mellitus, and appetite. For the simulated CKD dataset, key predictors included blood glucose random, blood urea, serum creatinine, potassium, hemoglobin, packed cell volume, specific gravity, albumin, red blood cells, pus cell clumps, appetite, pedal edema, and anemia. Using these selected features, the stacking ensemble achieved 100.0% accuracy on the UCI CKD dataset and 96.7% accuracy on the simulated dataset, both with negligible misclassification rates. Bootstrap confidence intervals confirmed the robustness of these results. The findings highlight that combining systematic feature selection with stacking significantly improves predictive accuracy while maintaining interpretability. This integrated framework offers a reliable tool for early CKD detection and can support clinical decision-making in real-world healthcare settings. Future work will focus on expanding validation across multi-site data and developing a clinical decision support interface.

Keywords: Chronic Kidney Disease (CKD), predisposing risk factors, stacking machine learning algorithm, disease prediction, feature selection.

AMS Classification: 62H30, 62P10, 62F40.

1. Introduction

Chronic kidney disease (CKD) is a global health crisis, affecting over 850 million people, with prevalence rates ranging from 8% to 16% [1][2][3]. Known as the "silent epidemic," CKD contributes to 1.2 million deaths annually and is projected to become the fifth leading cause of death by 2040 [1][2][17]. The burden is particularly severe in low- and middle-income countries, where limited access to diagnosis, healthcare infrastructure, and socioeconomic factors worsen its impact [4][5][20][23]. Due to its asymptomatic early stages, CKD often goes undiagnosed until significant kidney damage occurs, reducing treatment effectiveness [6][11]. Risk factors like diabetes, hypertension, cardiovascular diseases, smoking, and obesity are key contributors to CKD development and progression [6][9][15][16]. Current diagnostic methods, based on glomerular filtration rate (GFR) and albuminuria, often fail to identify high-risk individuals early [2][8][33]. As CKD advances, it causes severe complications, such as hypertension, anemia, and bone disease, requiring costly interventions like dialysis or transplantation [7][29][30][31]. Alarmingly, 96% of CKD patients remain unaware of their condition, highlighting the need for better predictive models [6][11]. The diagnostic framework for CKD involves GFR staging (G1–G5) and albuminuria classification (A1–A3), but these markers are insufficient for early detection. Figure 1 illustrates the risk stratification framework recommended by the Kidney Disease: Improving Global Outcomes (KDIGO) guidelines [33]. Advanced machine learning (ML) models can improve early detection by uncovering hidden patterns in clinical datasets, enhancing prediction accuracy [9][12][14][21][24]. While traditional ML models show promise, challenges like class imbalance, missing data, and feature redundancy often reduce their effectiveness [5][11][13][20]. Effective preprocessing techniques, such as feature selection and class balancing (e.g., SMOTE), are necessary to improve model accuracy [14][19][25][27]. This study aims to identify key predisposing factors for CKD and optimize prediction using a stacking machine learning algorithm. Two datasets are used: a real-world dataset with 400 instances and 24 features from the UCI ML repository [34] and a synthetic dataset with 1,000 samples to validate model robustness and generalizability.

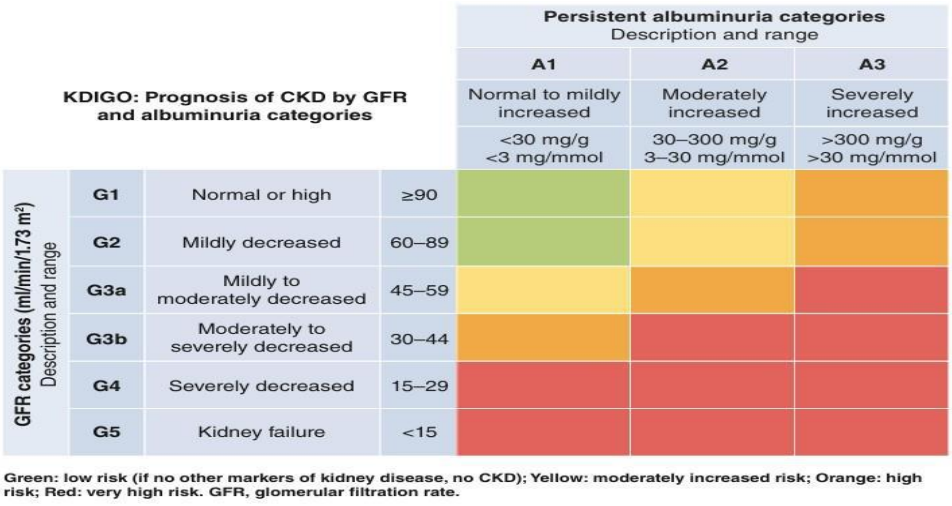

Figure 1: Prognosis of CKD by GFR and albuminuria categories (Source: KDIGO, 2024).

The study applies feature selection and SMOTE to enhance model performance. The stacking machine learning algorithm, which combines multiple models to improve accuracy, is used to assess scalability and reliability for clinical applications. By integrating advanced ML methods, this research aims to develop a robust, scalable CKD prediction framework to improve early detection and healthcare interventions [4][10][13][26][28][32].

The structure of this paper is organized to guide the reader through a systematic exploration of the research. Section 2 reviews existing literature, synthesizing past studies and the evolution of CKD predictive models. Section 3 covers dataset description, preprocessing steps, feature selection techniques, and the stacking ensemble method. Section 4 presents and analyzes the results from both the UCI CKD and simulated CKD datasets, including accuracy comparisons and benchmarking. Finally, section 5 concludes the study.

2. Related Work

CKD prediction has garnered significant attention due to its critical role in early diagnosis and patient management. Numerous studies have explored machine learning (ML) and deep learning (DL) techniques for CKD prediction, with a focus on feature selection strategies to improve accuracy. Poonia et al. (2022) [3] developed a feature-based predictive model using Logistic Regression (LR), K-Nearest Neighbors (KNN), Artificial Neural Networks (ANN), and others, with LR achieving 98.75% accuracy when combined with Chi-Square feature selection. Similarly, Muntasir Nishat et al. (2021) [4] demonstrated that Random Forest (RF) outperformed other models, reaching 99.75% accuracy. Krishnamurthy et al. (2021) [5] showed that Convolutional Neural Networks (CNN) were effective for CKD progression, achieving 89% accuracy for 6-month predictions. Ensemble methods have also proven effective. Jongbo et al. (2020) [6] achieved perfect accuracy (100%) using Random Subspace with KNN. Ilyas et al. (2021) [8] found J48 to be the top performer for stage-wise CKD classification. Other studies, including Senan et al. (2021) [9] and Khan et al. (2020) [10], reported high accuracy using Random Forest, with 100% performance in some cases. Deep learning techniques like ANN, LSTM, and GRU have been increasingly employed. Akter et al. (2021) [19] achieved 99% accuracy using ANN, while other studies, such as Dutta et al. (2024) [14], reported Logistic Regression achieving 100% accuracy. Despite the advancements, challenges remain, including the need for optimal feature selection and handling class imbalance. While ensemble and hybrid models have shown promise, their real-world applicability remains underexplored. This study aims to address these gaps by identifying predisposing risk factors for CKD and optimizing disease prediction using a stacking machine learning algorithm, integrating feature selection and class balancing techniques to improve prediction accuracy.

3. Materials and Methodology

3.1 Dataset Description, Preprocessing, and Simulation Study

Robust and reliable predictive modeling requires datasets that are both representative of real-world clinical conditions and suitable for computational analysis. To this end, our study employs a dual-dataset framework: the widely used Chronic Kidney Disease (CKD) dataset from the UCI Machine Learning Repository [34], and a synthetic CKD dataset generated through a carefully designed simulation study. The UCI dataset provides a benchmark for comparability with prior research, while the simulated dataset expands the scale and diversity of the data, enabling rigorous evaluation of model robustness and generalizability.

3.1.1 UCI CKD Dataset

The cornerstone of this study is the Chronic Kidney Disease dataset from the UCI Machine Learning Repository [34], a widely recognized benchmark in biomedical machine learning. It comprises 400 patient records, each annotated with 24 attributes spanning demographic, clinical, and laboratory variables, along with a binary target variable that classifies patients as CKD-positive or CKD-negative. To improve clarity and align with clinical reporting standards, abbreviated feature names in the original dataset (e.g., bp, sg) were expanded into their full clinical descriptors (blood pressure, specific gravity). This refinement facilitates interpretation by both healthcare professionals and computational scientists. A complete overview of the dataset including feature roles, data types, units of measurement, and missing-value counts is provided in Table 1, which forms the foundation for subsequent modeling.

Table 1: Description of the UCI Chronic Kidney Disease (CKD) dataset.

Attribute	Role	Data Type	Description	Units	Missing Values
age	Feature	Numerical (Integer)	Age of the patient	Years	9
blood pressure	Feature	Numerical (Integer)	Blood pressure	mmHg	12
specific gravity	Feature	Categorical	Urine specific gravity	-	47
albumin	Feature	Categorical	Albumin levels in urine	-	46
sugar	Feature	Categorical	Sugar levels in urine	-	49
red blood cells	Feature	Binary (Nominal)	Presence of red blood cells	Present/Absent	152
pus cell	Feature	Binary (Nominal)	Presence of pus cells	Present/Absent	65
pus cell clumps	Feature	Binary (Nominal)	Presence of pus cell clumps	Present/Absent	4
bacteria	Feature	Binary (Nominal)	Presence of bacteria	Present/Absent	4
blood glucose random	Feature	Numerical (Integer)	Random blood glucose levels	mg/dL	44
blood urea	Feature	Numerical (Integer)	Blood urea levels	mg/dL	19
serum creatinine	Feature	Numerical (Float)	Serum creatinine levels	mg/dL	17
sodium	Feature	Numerical (Integer)	Sodium levels	mEq/L	87
potassium	Feature	Numerical (Float)	Potassium levels	mEq/L	88
hemoglobin	Feature	Numerical (Float)	Hemoglobin levels	g/dL	52

packed cell volume	Feature	Numerical (Integer)	Packed cell volume	-	71
white blood cell count	Feature	Numerical (Integer)	White blood cell count	cells/cmm	106
red blood cell count	Feature	Numerical (Float)	Red blood cell count	millions/cmm	131
hypertension	Feature	Binary (Nominal)	Hypertension status	Yes/No	2
diabetes mellitus	Feature	Binary (Nominal)	Diabetes mellitus status	Yes/No	2
coronary artery disease	Feature	Binary (Nominal)	Coronary artery disease status	Yes/No	2
appetite	Feature	Binary (Nominal)	Appetite level	Good/Poor	1
pedal edema	Feature	Binary (Nominal)	Presence of pedal edema	Yes/No	1
anemia	Feature	Binary (Nominal)	Presence of anemia	Yes/No	1
class	Target Variable	Binary (Nominal)	Classification of CKD or Not CKD	CKD/Not CKD	0

3.1.2 Data Preprocessing and EDA

A rigorous, leakage-safe preprocessing and exploratory data analysis (EDA) pipeline was developed to ensure clinical reliability and statistical integrity prior to model development. The UCI Chronic Kidney Disease (CKD) dataset exhibited challenges typical of clinical records such as missing values, inconsistent categorical labels, numeric anomalies, and moderate class imbalance. To address these, we implemented a structured workflow covering data cleaning, encoding, transformation, and balancing. All variable names were standardized, and inconsistent textual entries (e.g., variants of “yes/no” and tabbed strings) were harmonized. Mixed-type numeric fields were coerced to valid numeric formats. Missing values were imputed using median (continuous) and mode (categorical) strategies. Binary clinical attributes were encoded via clinically meaningful mappings (normal = 0, abnormal = 1; no = 0, yes = 1; appetite: good = 0, poor = 1). The target variable was encoded as CKD = 1 and not-CKD = 0, ensuring consistent interpretation across analyses. To handle extreme numeric deviations without discarding patients, we applied Z-score capping ($|Z| > 3$) using statistics learned only from the training subset to avoid information leakage; values were clipped to clinically plausible ranges. Continuous variables were then normalized with Min–Max scaling to [0, 1] to equalize feature influence. Data were stratified 70%/30% (train/test) while preserving the original class distribution (CKD = 62.5%, not-CKD = 37.5%). To mitigate imbalance, SMOTE was applied only to the training data, yielding a balanced training set (175 CKD vs. 175 not-CKD). Figure 2 illustrates the pre-SMOTE training distribution, the overall raw ratio, and the balanced training distribution after resampling, demonstrating unbiased learning without contaminating the held-out test set.

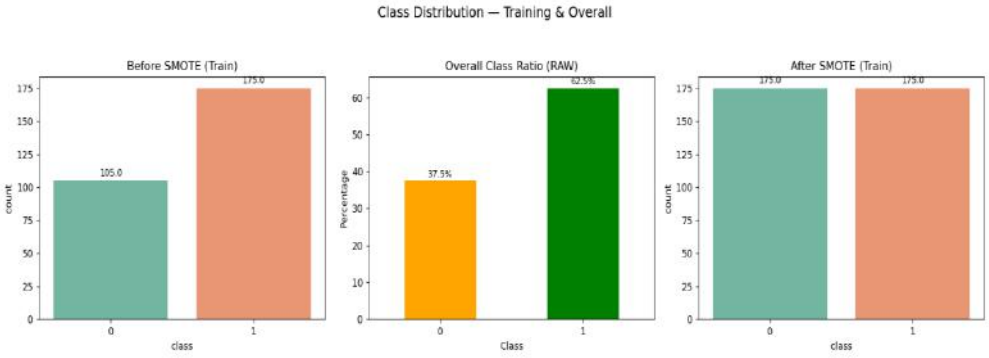

Figure 2: Class distribution analysis.

A correlation heatmap in Figure 3 revealed strong negative associations between CKD (coded 1) and key renal-function indicators such as hemoglobin ($r = -0.77$), packed cell volume ($r = -0.74$), and specific gravity ($r = -0.73$) consistent with anemia and impaired urine concentration in CKD. Positive associations were observed for albumin ($r = 0.63$), blood glucose random ($r = 0.42$), and blood urea ($r = 0.38$), reflecting metabolic stress and reduced filtration. Physiologically coherent relationships (e.g., blood urea–serum creatinine ≈ 0.59; hemoglobin–packed cell volume ≈ 0.90) further validated internal consistency. A Random Forest feature-importance analysis shown in Figure 4 identified hemoglobin, packed cell volume, serum creatinine, specific gravity, and red blood cell count as the most influential predictors, followed by comorbidity indicators (hypertension, diabetes mellitus). In contrast, coronary artery disease, bacteria, and pus cell clumps contributed minimally.

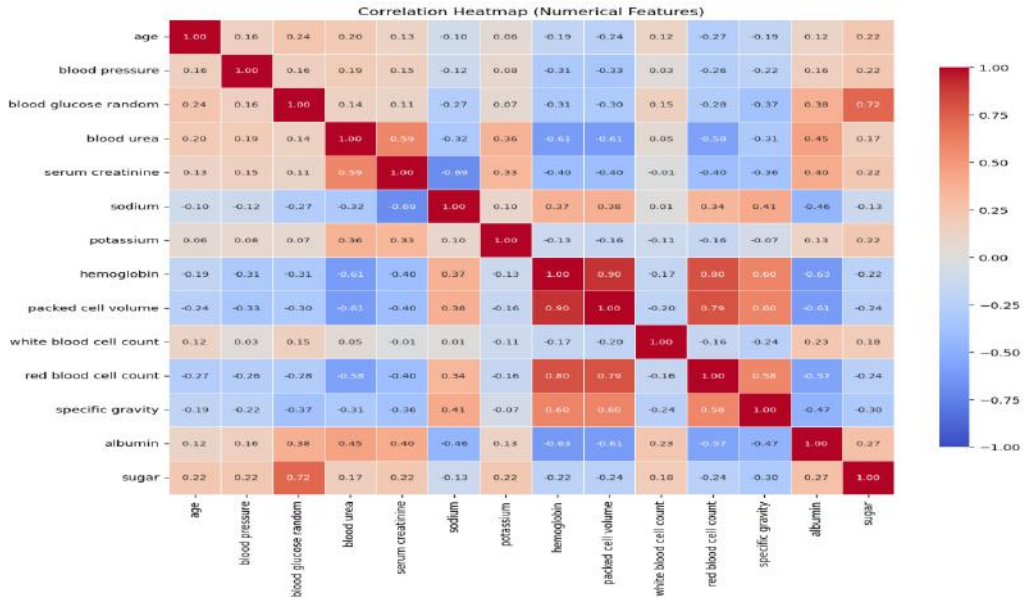

Figure 3: Correlation heatmap of numerical features.

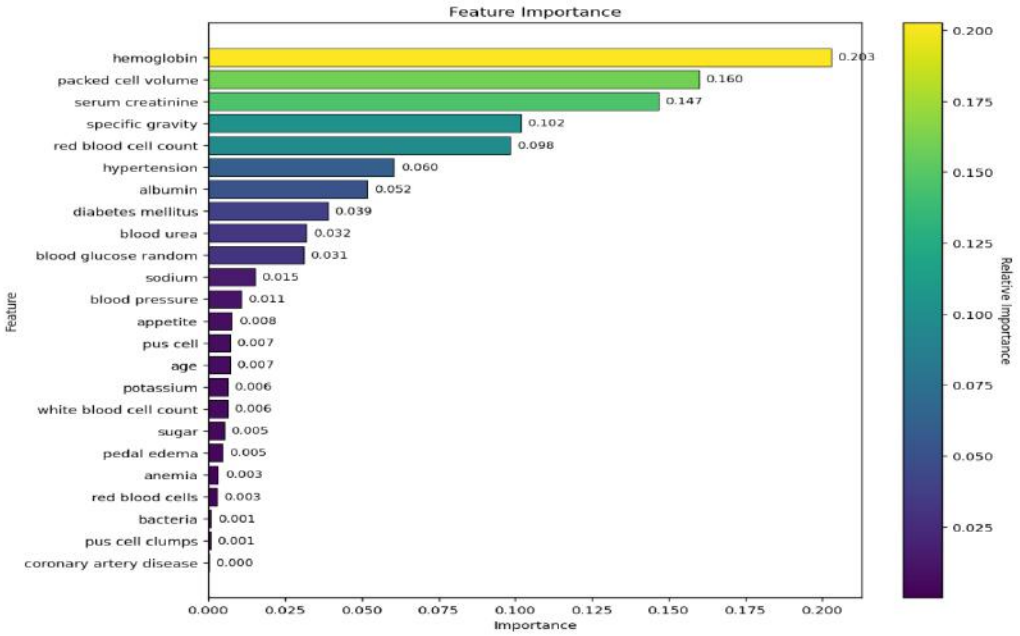

Figure 4: Feature importance plot.

Together, these steps produced a clinically coherent, statistically balanced, and leakage-free dataset, forming a robust foundation for subsequent feature selection and ensemble modeling.

3.1.3 Simulation Data Generation

To complement the UCI CKD dataset and evaluate model robustness under controlled yet realistic conditions, a synthetic CKD cohort of 1,000 patients was generated. The simulator reproduced the same 24 clinical attributes and target variable as the UCI dataset, ensuring direct comparability while introducing realistic variability, inter-feature correlations, and missingness patterns. All random processes were seeded (R = 42) for reproducibility, and the target variable was encoded as ckd = 1 and not ckd = 0.

A hierarchical data-generation framework was implemented, consisting of five major components:

1. Demographic and Comorbidity Modeling: Patient age was sampled uniformly between 20 and 80 years. The probability of diabetes mellitus (DM) increased with age according to a logistic function, while blood pressure (BP) was modeled as $N\{115 + 0.4(Age - 50) + 8 * DM, 12^2\}$, truncated between 80 and 200 mmHg. Hypertension (HTN) was stochastically derived from BP via a logistic relationship, producing approximately 50% prevalence near 133 mmHg.

2. CKD Risk and Outcome Generation: CKD probability was determined using a latent logistic model calibrated to achieve an overall prevalence of ~55%:

$$logit(p_{CKD}) = 0.02(Age - 50) + 1.05DM + 0.95HTN + c \quad (1)$$

where the intercept c was iteratively adjusted to match the target prevalence.

3. Class-Conditional Laboratory Profiles: Six key biochemical variables such as blood glucose random (BGR), blood urea, serum creatinine, sodium, potassium, and hemoglobin were generated from class-specific multivariate Gaussian distributions, ensuring realistic inter-variable correlations through Cholesky decomposition. Table 2 shows parameters for class-conditional laboratory generation.

Table 2: Parameters for class-conditional laboratory generation.

Variable	μ (Non-CKD)	μ (CKD)	σ (Non-CKD)	σ (CKD)	Bounds (range)
Blood glucose random (mg/dL)	120	185	25	35	[60, 450]
Blood urea (mg/dL)	35	65	12	18	[5, 220]
Serum creatinine (mg/dL)	0.9	1.4	0.25	0.40	[0.5, 12]
Sodium (mEq/L)	140	138	3.0	3.5	[125, 150]
Potassium (mEq/L)	4.2	4.5	0.35	0.40	[3.0, 6.5]
Hemoglobin (g/dL)	13.5	10.2	1.0	1.2	[5, 18]

4. Derived Hematologic and Categorical Features: Hematologic dependencies were enforced such that PCV≈3×Hb (bounded [20, 60]) and lower red blood cell counts were assigned to anemic subjects (Hb<11). Urinalysis attributes (specific gravity, albumin, sugar) and clinical indicators (microscopy, appetite, edema, coronary artery disease) were sampled from class-conditional multinomial distributions that matched CKD-related clinical prevalence patterns.

5. Missingness Mechanisms: To replicate real clinical incompleteness, feature-specific missing rates (up to 60%) were assigned under three mechanisms: (i) missing completely at random (MCAR) random omissions, (ii) missing at random (MAR) reduced missingness for CKD patients in key lab tests, (iii) missing not at random (MNAR) value-dependent omissions, such as high albumin being less likely to be missing.

The resulting dataset preserved the exact schema of the UCI CKD data, with all values constrained within physiologically valid ranges. Randomness and correlations were clinically informed rather than arbitrary. No distributional drift was applied in this study. Overall, the simulated dataset is clinically coherent, statistically realistic, and structurally aligned with the UCI CKD dataset, providing a robust foundation for model validation under controlled variability and uncertainty.

3.2 Feature Selection

Feature selection is a crucial preprocessing step that identifies the most relevant features in a dataset, improving model interpretability, reducing complexity, and minimizing overfitting. In this study, various feature selection techniques were applied to identify key predictors for classifying CKD, enhancing model efficiency and accuracy while reducing dimensionality. The following methods were applied:

3.2.1 Logistic Regression (LR) Feature Selection

Logistic Regression (LR) is a linear model commonly used for binary classification. It predicts the probability of a binary outcome using a linear combination of input features transformed by the sigmoid function. In feature selection, LR leverages the interpretability of its coefficients, where

the magnitude of each coefficient $|\beta_i|$ indicates the importance of the corresponding feature. The relationship between the independent variables (X) and the target variable (y) is modeled as:

$$P\,(1 \mid X) = \frac{1}{1+ e^{-(\beta_0+\sum_{i=1}^{n}\beta_i X_i)}} \quad (2)$$

To ensure comparability across features, standardization was applied:

$$X_{scaled} = \frac{X-\mu}{\sigma} \quad (3)$$

where μ is the mean and σ is the standard deviation of each feature. Using the SMOTE-balanced training dataset, LR was trained with a maximum iteration limit of 10,000. Features were ranked by their absolute coefficients $|\beta_i|$, and the top 10 features were selected.

3.2.2 Recursive Feature Elimination (RFE) Feature Selection

Recursive Feature Elimination (RFE) is a wrapper-based method that selects features by recursively training a model, ranking features based on their importance, and removing the least significant ones until the desired number of features is reached. In RFE, feature importance is evaluated iteratively. For linear models, importance is typically based on the magnitude of coefficients $|\beta_i|$, while for tree-based models, it may rely on metrics like Gini importance. The RFE process involves (i) Training the model on the current feature set, (ii) Computing feature importance $I(X_i)$:

$$I(X_i) = |\beta_i| \text{ (linear models) or } I(X_i) = Gini\ importance \text{ (tree models)}, \quad (4)$$

(iii) Eliminating the least important feature and repeating until the desired number of features remains. Standardization was performed to ensure uniform scales:

$$X_{scaled} = \frac{X-\mu}{\sigma} \quad (5)$$

For CKD prediction, RFE was applied with Logistic Regression as the base estimator to select the top 10 features.

3.2.3 Random Forest (RF) Feature Selection

Random Forest (RF) is an ensemble learning method based on decision trees. It computes feature importance by measuring the decrease in impurity (e.g., Gini index) when a feature is included in the model. RF ranks features by aggregating their importance scores across all decision trees in the ensemble. Feature importance is calculated as:

$$Feature\ importance\ (RF) = \frac{\sum_{t=1}^{T}\Delta\, I_t(f)}{T} \quad (6)$$

where T represents total number of trees and $\Delta\, I_t(f)$ represents reduction in impurity for feature f in tree t. Although RF does not require feature scaling, standardization was applied for consistency:

$$X_{scaled} = \frac{X-\mu}{\sigma} \quad (7)$$

Using 100 estimators, the top 10 features with the highest importance scores were selected.

3.2.4 Mutual Information (MI) Feature Selection

Mutual Information (MI) is a measure of dependency between two variables. It captures both linear and non-linear relationships by quantifying the reduction in uncertainty of one variable given knowledge of the other. For feature selection, MI evaluates how much information each feature (X_i) contributes to predicting the target variable (y):

$$I(X_i; y) = \sum_{x \in X_i}\sum_{y \in Y} P(x,y)\log\left(\frac{P(x,y)}{P(x)\,P(y)}\right) \quad (8)$$

where $P(x, y)$ is the joint probability of X_i and y, $P(x)$ is the marginal probability of X_i and $P(y)$ is the marginal probability of y. Although MI does not require scaling, features were standardized for consistency:

$$X_{scaled} = \frac{X-\mu}{\sigma} \quad (9)$$

The top 10 features based on MI scores were selected.

3.2.5 Chi-Square (χ^2) Feature Selection

The Chi-Square (χ^2) test evaluates the independence between categorical features and the target variable. It assesses whether observed frequencies differ significantly from expected frequencies. For each feature (X_i), the χ^2 statistic is calculated as:

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i} \quad (10)$$

where O_i is the observed frequency and E_i is the expected frequency. Features were standardized and transformed to non-negative values to satisfy test assumptions:

$$X_{scaled} = \frac{X-\mu}{\sigma}, \ |X_{scaled}| \quad (11)$$

Using the SelectKBest class with the χ^2 scoring function, the top 10 features were selected.

3.2.6 Principal Component Analysis (PCA) for Feature Selection

Principal Component Analysis (PCA) is a dimensionality reduction technique that transforms the original feature set into a new orthogonal basis, called principal components, which maximize variance. Instead of retaining principal components, features were ranked based on their contributions (loadings) to the principal components. Each principal component (PC_k) is defined as:

$$PC_k = \sum_{i=1}^{k} w_{ik} X_i \quad (12)$$

where w_{ik} is the loading of the $i-th$ feature in the $k-th$ component and X_i is the original feature value. Feature importance was determined by summing the absolute loadings across all components:

$$Feature\ importance\ (PCA) = \sum_{k=1}^{m} |w_{ik}| \quad (13)$$

Features were standardized before applying PCA:

$$X_{scaled} = \frac{X-\mu}{\sigma} \quad (14)$$

The top 10 features with the highest cumulative loadings were selected.

3.2.7 Common Features Selection

Let $X \in \mathbb{R}^{n \times d}$ represent the original dataset, where n is the number of samples, d is the total number of features, and Y denotes the corresponding target variable. We consider k feature selection methods $F_1, F_2, \ldots, F_k$, where each method F_i selects a subset of features $S_i \subseteq X$. The ensemble-selected feature set $S_{ensemble}$ is defined as the intersection of the subsets identified by all k methods:

$$S_{ensemble} = \bigcap_{i=1}^{k} S_i = S_1 \cap S_2 \cap \cdots \cap S_k \quad (15)$$

In practice, the ensemble selection is determined using a majority-vote approach. Specifically, a feature is considered common if it is selected by at least a predefined number of methods. The final common feature set is denoted by $|S_{ensemble}|$, which represents the number of features that appear in at least 3 methods (in the case of a threshold of 3).

In this study, this methodology across six feature selection techniques (LR, RFE, RF, MI, χ^2, and PCA) identified 10 common features for UCI CKD dataset and 13 common features for simulated CKD dataset, critical for CKD prediction. Table 3 lists the common features identified across all methods for both datasets.

Table 3: Common features selected across methods (majority vote)

UCI CKD Dataset	Simulated CKD Dataset
serum creatinine, hemoglobin, packed cell volume, red blood cell count, specific gravity, albumin, sugar, hypertension, diabetes mellitus, appetite	blood glucose random, blood urea, serum creatinine, potassium, hemoglobin, packed cell volume, specific gravity, albumin, red blood cells, pus cell clumps, appetite, pedal edema, anemia

3.3 Stacking Ensemble

The stacking ensemble method stands as an advanced machine learning technique, adept at synthesizing the strengths of diverse predictive models to achieve enhanced accuracy and generalization. Unlike conventional ensemble approaches such as bagging and boosting, stacking thrives on leveraging the heterogeneity of its base models. By combining their complementary capabilities through a strategically trained meta-learner, it delivers predictions that are both precise and robust. In this study, the stacking ensemble approach is harnessed to elevate the classification performance for CKD. The process begins with a dataset $X \in \mathbb{R}^{n \times d}$, where n represents the number of samples, d is the total number of features, and Y denotes the corresponding target variable. A set of k base learners $h_1, h_2, \ldots, h_k$ is trained, each independently learning a mapping $h_i : X \rightarrow \hat{Y}_{h_i}$. These models generate predictions $\hat{Y}_{h_i} = h_i(X), i = 1, 2, \ldots, k,$ which form the columns of a new dataset $Z \in \mathbb{R}^{n \times k}$. Here, each column of Z encapsulates the output of a specific base model. The meta-learner h_{meta} is then trained on Z to establish a higher-order relationship between the base learners' predictions and the true target Y. Mathematically, this relationship is represented as $h_{meta} : Z \rightarrow \hat{Y}_{meta}$. The final ensemble prediction is computed as $\hat{Y}_{stacking} = h_{meta}(Z)$.

3.4 Proposed Methodology

The proposed chronic kidney disease (CKD) prediction framework, shown in Figure 5, follows a structured pipeline with four main phases: data preprocessing, feature selection, model training, and evaluation. The goal of this approach is to build a reliable and interpretable model that can predict CKD while avoiding data leakage and overfitting. In the data preprocessing phase, the dataset was cleaned and standardized to ensure consistency. All column names were unified, and inconsistent categorical entries such as tab-delimited or mixed-case "yes/no" responses were corrected. Numeric columns stored as text were converted to proper numerical types. Missing values were filled using the median for continuous variables and the mode for categorical variables, maintaining the overall distribution of each feature. Binary clinical variables were encoded using simple, clinically meaningful mappings (for example, normal/good/no = 0 and abnormal/poor/yes = 1), and the target variable was labeled as CKD = 1 and not-CKD = 0. To handle outliers without removing any records, values beyond three standard deviations from the mean were adjusted using Z-score capping ($|Z| > 3$), applied only to the training set to prevent information leakage. Continuous variables were then normalized with Min–Max scaling to bring all features into the [0, 1] range and prevent features with large numeric ranges from dominating the model. The dataset was split into 70% training and 30% testing subsets while keeping the same

class proportions. Because the data showed a moderate class imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) was used only on the training data to generate synthetic minority samples. This produced a balanced training set and ensured that oversampling did not influence the test results. In the feature selection phase, six different methods were used to identify the most relevant predictors: Logistic Regression coefficients, Recursive Feature Elimination (RFE), Random Forest feature importance, Mutual Information, Chi-square (χ^2) tests, and Principal Component Analysis (PCA). Each method captured different statistical or structural relationships among the variables. To ensure stability, features that appeared in at least three of these methods were selected. This majority-vote approach reduced redundancy and kept the most informative and clinically meaningful variables. During the modeling phase, several machine learning algorithms were trained and compared, including K-Nearest Neighbors (KNN), Support Vector Machine with an RBF kernel (SVM-RBF), Random Forest (RF), Gaussian Naïve Bayes (GNB), Logistic Regression (LR), Multi-Layer Perceptron (MLP), and Extreme Gradient Boosting (XGBoost). Each model was trained in a leakage-safe pipeline where scaling, sampling, and parameter tuning were performed only on the training data. To combine the strengths of individual models, a stacking ensemble was built. The base models (level 0) generated probabilistic predictions that were then combined by a Logistic Regression meta-learner (level 1) to produce the final output. This ensemble approach improved generalization and reduced the variance of individual classifiers. In the evaluation phase, model performance was measured using Accuracy, Precision, Recall, F1-Score, and Misclassification Rate (MCR) for both training and testing datasets. These metrics allowed comparison of model fit and generalization performance. To assess the statistical reliability of each metric, bootstrap resampling (2,000 iterations) was used to compute 95% confidence intervals (CIs). This provided a more stable estimate of model performance under sampling variability. Overall, the proposed methodology brings together careful preprocessing, consistent feature selection, diverse modeling techniques, and statistically sound evaluation. It provides a clear and reproducible framework for predicting CKD with high accuracy while maintaining clinical interpretability and methodological transparency.

3.5 Model Evaluation Metrics and Protocols

The performance of predictive models was evaluated using the following metrics:

Accuracy: Measures the proportion of correct predictions:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{16}$$

where, TP (True Positives), TN (True Negatives), FP (False Positives), and FN (False Negatives) are the components of the confusion matrix.

Precision: Indicates the accuracy of positive predictions:

$$Precision = \frac{TP}{TP+FP} \tag{17}$$

Recall (Sensitivity): Assesses the ability to identify all positive instances:

$$Recall = \frac{TP}{TP+FN} \tag{18}$$

F1_Score: Balances precision and recall. It is formulated as:

$$F1\_Score = \frac{2 \times Precision \times Recall}{Precision+Recall} \tag{19}$$

Misclassification Rate (MCR): The proportion of incorrect predictions:

$$MCR = 1 - Accuracy = \frac{FP+FN}{TP+TN+FP+FN} \quad (20)$$

These metrics help assess the model's overall performance and robustness in CKD classification.

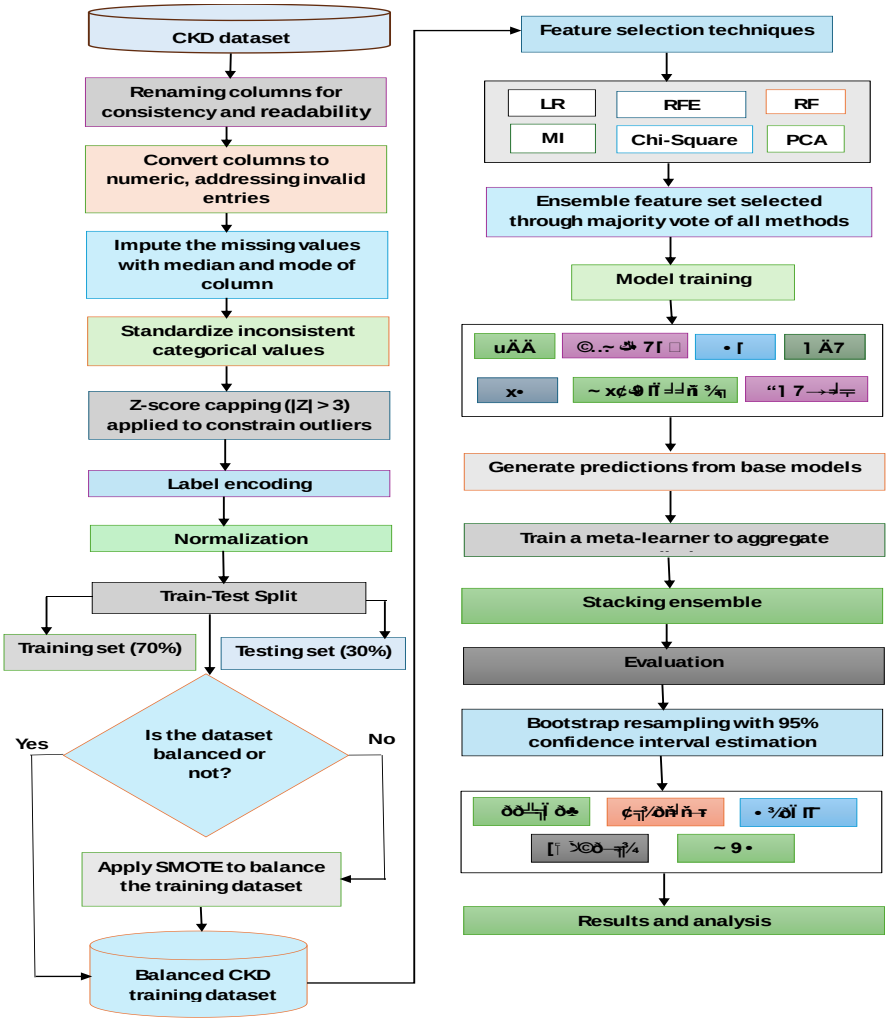

Figure 5: Methodology flowchart for chronic kidney disease prediction.

4. Analysis of the Results

This section presents a detailed evaluation of the proposed approach for identifying key CKD risk factors and improving predictive accuracy through the stacking ensemble model. The analysis includes results from the original UCI CKD dataset and the simulated dataset, covering model generalization, feature-based performance, and bootstrap confidence intervals for result stability.

4.1 Performance on UCI CKD Dataset

This subsection summarizes model performance using the real clinical dataset after full preprocessing and feature selection.

4.1.1 Comprehensive Algorithm Comparison (Train vs. Test)

Tables 4–8 and Figures 6–10 show the training and testing results for all implemented algorithms on the UCI CKD dataset. Overall, the models demonstrate high predictive performance with minimal generalization gaps, indicating excellent consistency between training and test outcomes. Stacking and XGBoost achieve 100.0% test accuracy, with Stacking matching its training score and XGBoost differing by only −0.4 percentage points. SVM (RBF), Random Forest, and MLP maintain 99.2% accuracy, while KNN achieves 98.3% and Logistic Regression 97.5%. Gaussian NB, though slightly lower at 95.0%, still performs reasonably well. Precision, recall, and F1-scores align closely with accuracy, confirming balanced detection of both CKD and non-CKD cases. Misclassification rates remain minimal 0.0% for Stacking and XGBoost, below 2.5% for the others, and 5.0% for Gaussian NB. Overall, the results confirm strong model generalization, with the Stacking Ensemble and XGBoost emerging as the most robust and reliable classifiers.

Table 4: Comprehensive algorithm comparison of accuracy.

Model	**Train Accuracy** (%)	**Test Accuracy** (%)
KNN	99.6%	98.3%
SVM (RBF)	100.0%	99.2%
Random Forest	100.0%	99.2%
Gaussian NB	95.7%	95.0%
Logistic Regression	100.0%	97.5%
MLP	100.0%	99.2%
XGBoost	99.6%	100.0%
Stacking Ensemble	**100.0%**	**100.0%**

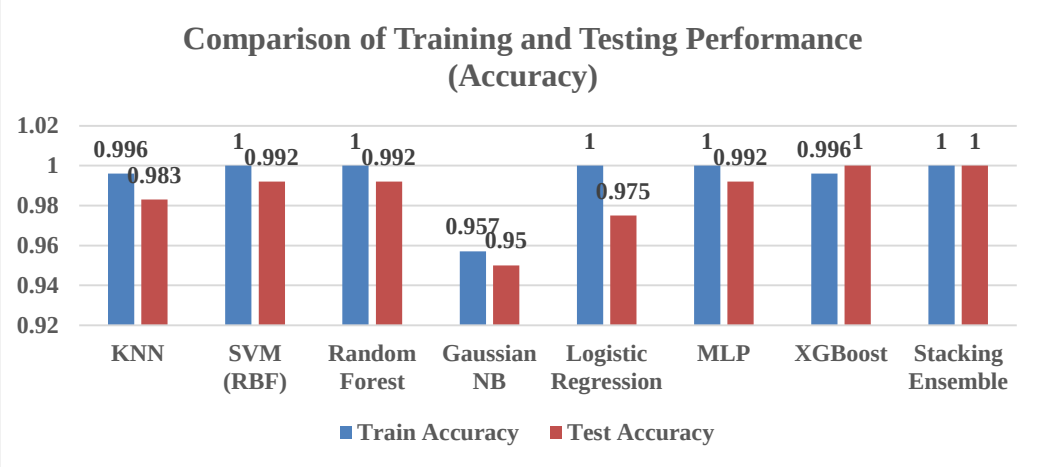

Figure 6: Comparison of training and testing performance (Accuracy).

Tables 5–8 and Figures 7–10 provide the corresponding comparisons for precision, recall, F1-score, and MCR, reinforcing that ensemble and kernel-based models consistently deliver the most stable and accurate results.

Table 5: Comprehensive algorithm comparison of precision.

Model	**Train Precision** (%)	**Test Precision** (%)
KNN	99.6%	98.4%
SVM (RBF)	100.0%	99.2%
Random Forest	100.0%	99.2%
Gaussian NB	96.2%	95.6%
Logistic Regression	100.0%	97.7%
MLP	100.0%	99.2%
XGBoost	99.6%	100.0%
Stacking Ensemble	**100.0%**	**100.0%**

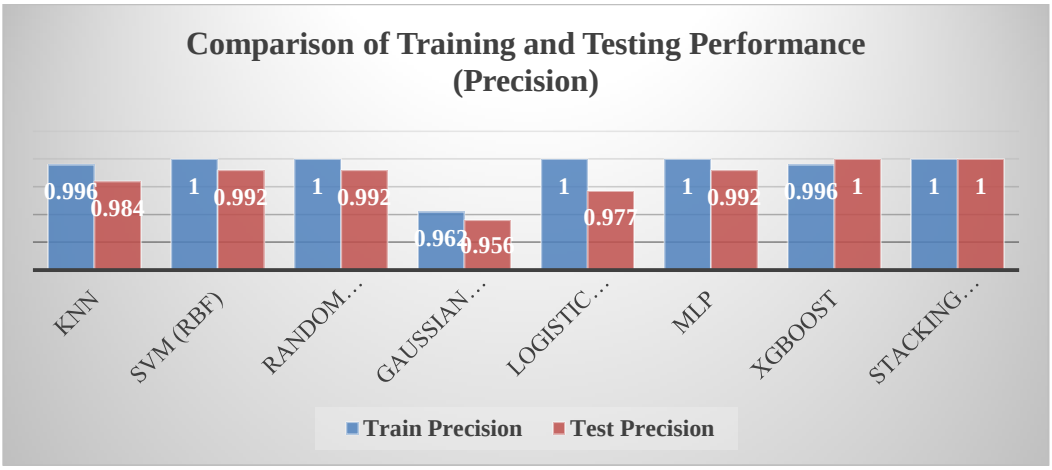

Figure 7: Comparison of training and testing performance (Precision).

Table 6: Comprehensive algorithm comparison of recall.

Model	**Train Recall** (%)	**Test Recall** (%)
KNN	99.6%	98.3%
SVM (RBF)	100.0%	99.2%
Random Forest	100.0%	99.2%
Gaussian NB	95.7%	95.0%
Logistic Regression	100.0%	97.5%
MLP	100.0%	99.2%
XGBoost	99.6%	100.0%
Stacking Ensemble	**100.0%**	**100.0%**

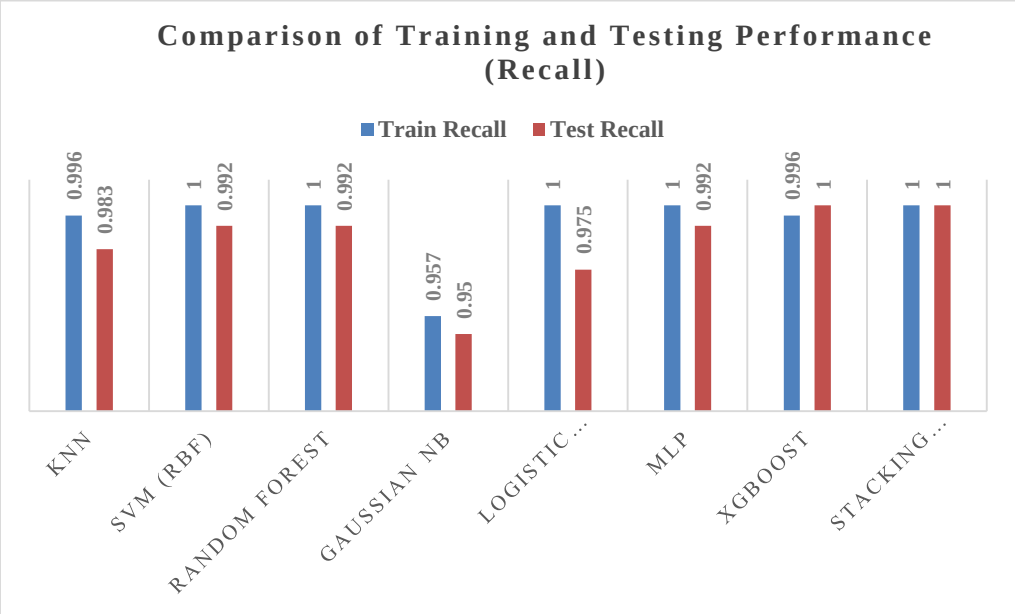

Figure 8: Comparison of training and testing performance (Recall).

Table 7: Comprehensive algorithm comparison of F1_Score.

Model	**Train F1_Score** (%)	**Test F1_Score** (%)
KNN	99.6%	98.3%
SVM (RBF)	100.0%	99.2%
Random Forest	100.0%	99.2%
Gaussian NB	95.8%	95.1%
Logistic Regression	100.0%	97.5%
MLP	100.0%	99.2%
XGBoost	99.6%	100.0%
Stacking Ensemble	**100.0%**	**100.0%**

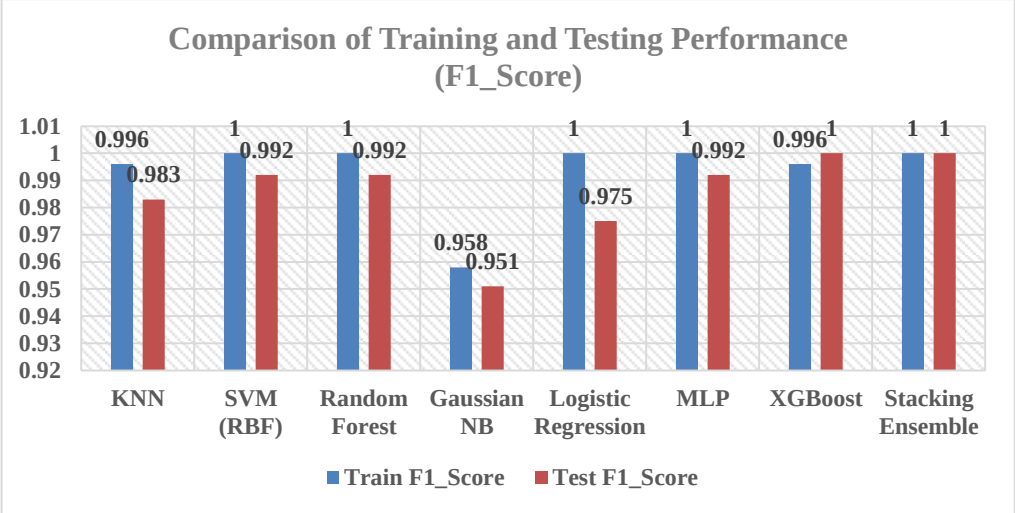

Figure 9: Comparison of training and testing performance (F1_Score).

Table 8: Comprehensive algorithm comparison of MCR.

Model	**Train MCR** (%)	**Test MCR** (%)
KNN	0.4%	1.7%
SVM (RBF)	0.0%	0.8%
Random Forest	0.0%	0.8%
Gaussian NB	4.3%	5.0%
Logistic Regression	0.0%	2.5%
MLP	0.0%	0.8%
XGBoost	0.4%	0.0%
Stacking Ensemble	**0.0%**	**0.0%**

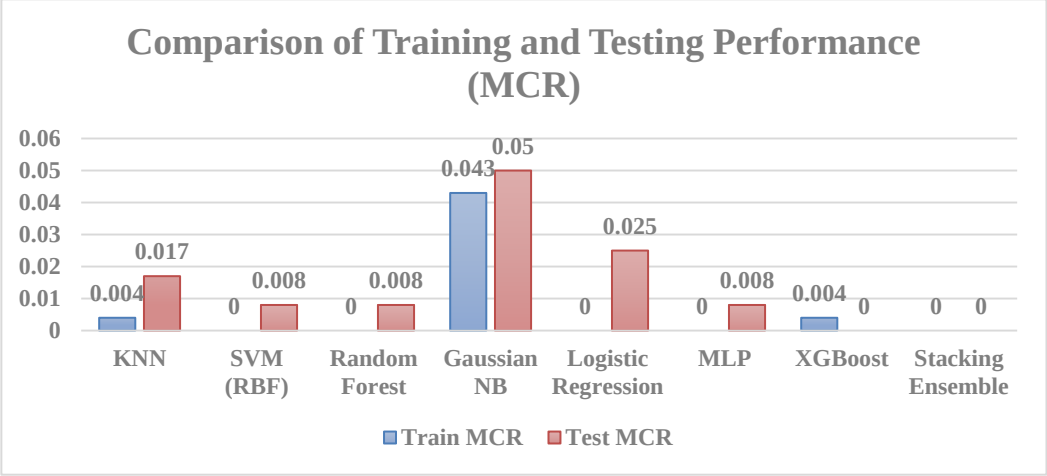

Figure 10: Comparison of training and testing performance (MCR).

4.1.2 Model Performance with Selected Features on the UCI CKD Dataset

After applying six complementary feature selection methods (LR, RFE, RF, MI, χ^2, and PCA), a refined subset of 10 clinically relevant features was identified as the most consistent predictors of CKD. Table 9 and Figures 11–15 show that this optimized feature set maintains or even enhances performance across all models. Stacking and XGBoost again reach 100.0% across all evaluation metrics, with zero misclassifications, confirming that feature optimization did not compromise model accuracy. SVM (RBF), Random Forest, and MLP sustain 99.2% performance with 0.8% MCR, while KNN and Logistic Regression remain strong at 98.3% and 97.5%, respectively. Gaussian NB continues to perform slightly below the others but maintains consistent predictive ability. These results demonstrate that feature selection successfully captured the most informative CKD predictors, improving interpretability while preserving accuracy. Ensemble-based models, especially the Stacking Ensemble, clearly show their advantage in combining multiple perspectives for more reliable classification.

Table 9: Performance metrics for models using selected features on the UCI CKD dataset.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score %	MCR (%)
KNN	98.3%	94.8%	98.3%	98.3%	1.7%
SVM (RBF)	99.2%	99.2%	99.2%	99.2%	0.8%
Random Forest	99.2%	99.2%	99.2%	99.2%	0.8%
Gaussian NB	95.0%	95.6%	95.0%	95.1%	5.0%
Logistic Regression	97.5%	97.7%	97.5%	97.5%	2.5%
MLP	99.2%	99.2%	99.2%	99.2%	0.8%
XGBoost	100.0%	100.0%	100.0%	100.0%	0.0%
Stacking Ensemble	**100.0%**	**100.0%**	**100.0%**	**100.0%**	**0.0%**

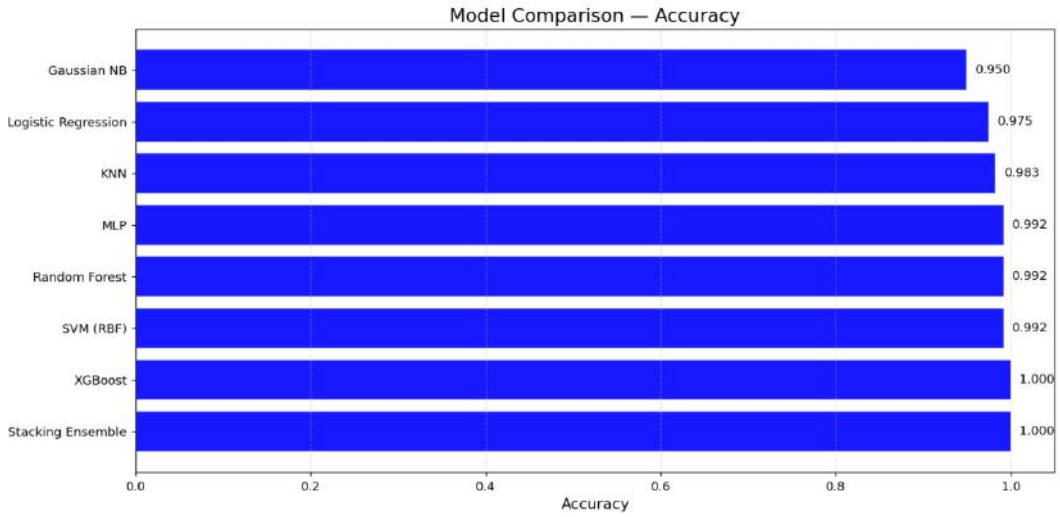

Figure 11: Comparative bar chart of model accuracy on the UCI CKD dataset.

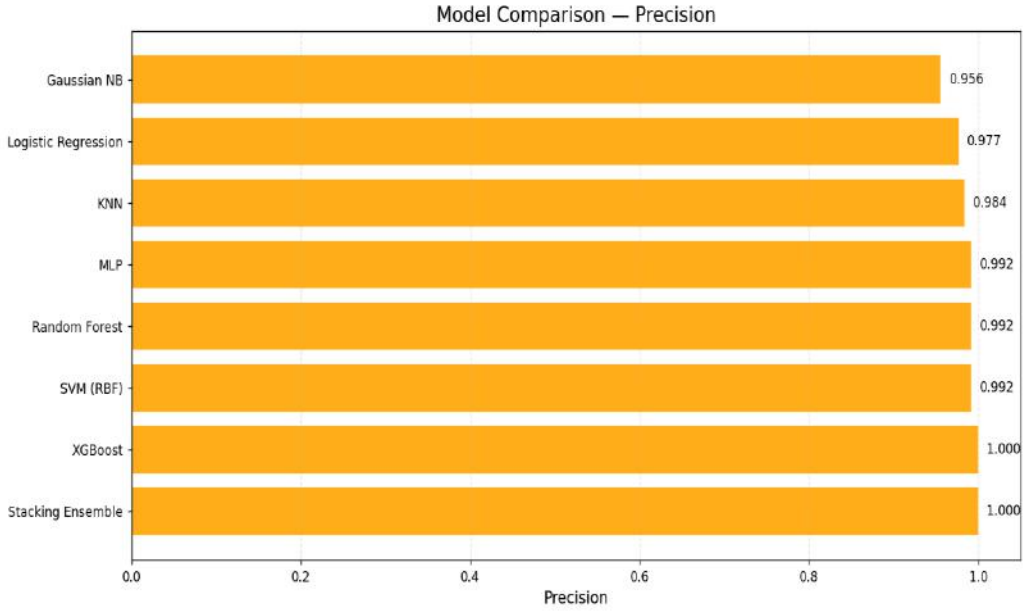

Figure 12: Comparative bar chart of model precision on the UCI CKD dataset.

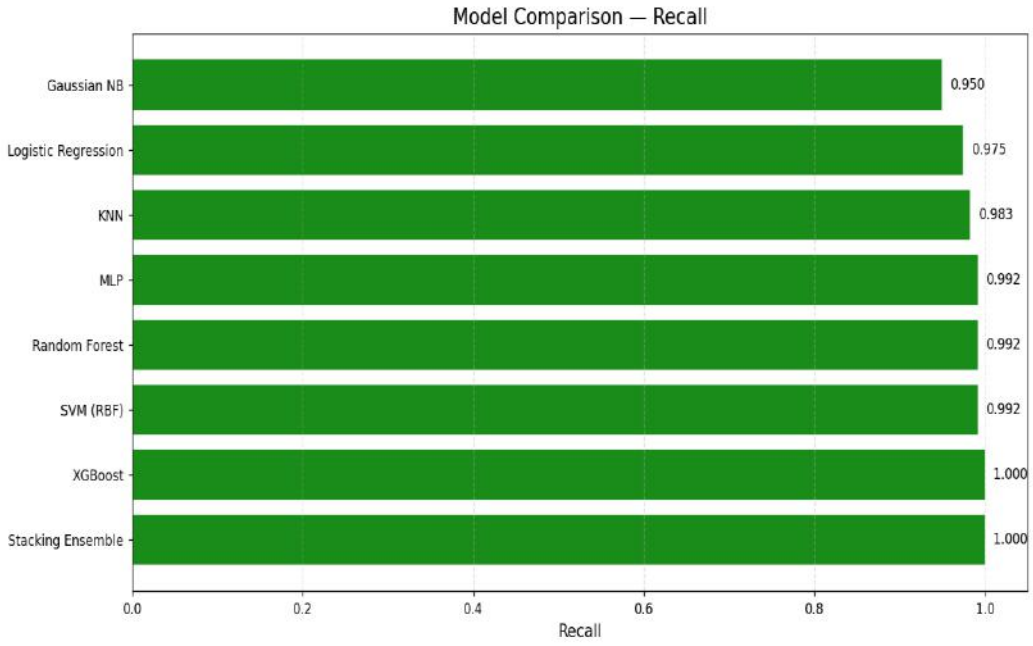

Figure 13: Comparative bar chart of model recall on the UCI CKD dataset.

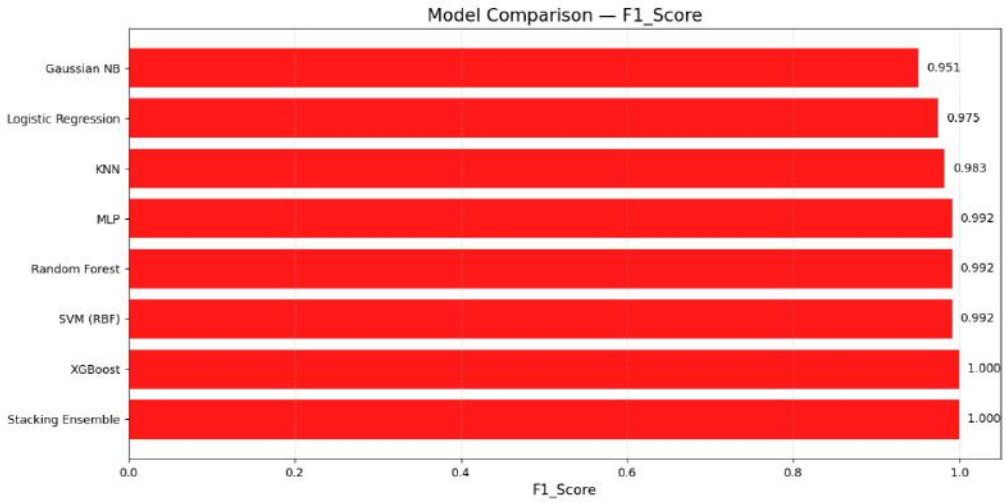

Figure 14: Comparative bar chart of model F1_Score on the UCI CKD dataset.

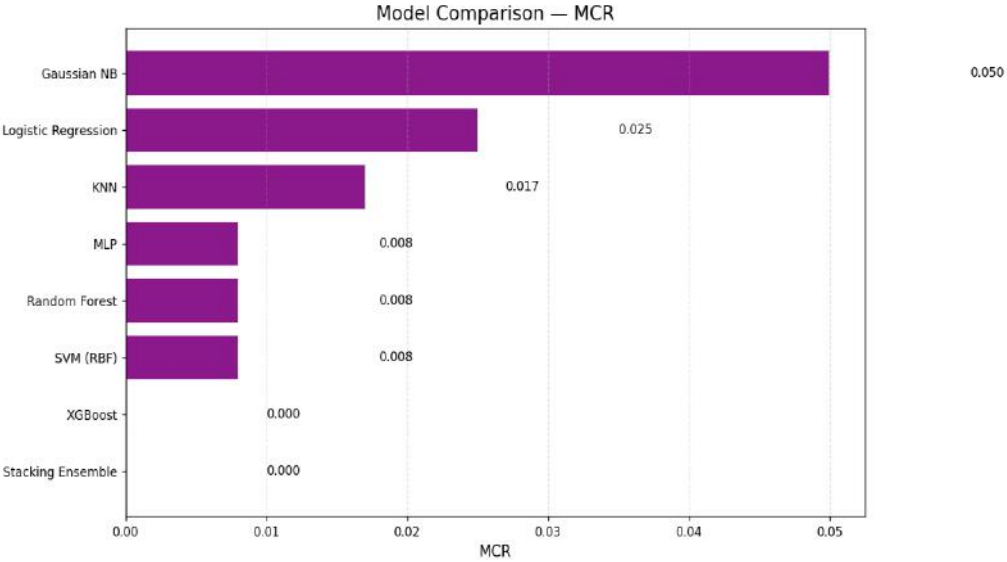

Figure 15: Comparative bar chart of model MCR on the UCI CKD dataset.

4.1.3 Bootstrap Performance with 95% Confidence Intervals on the UCI CKD Dataset

Figures 16–20 show the bootstrap-based 95% confidence intervals for all five metrics on the UCI CKD dataset. Stacking and XGBoost again produce perfect stability, with intervals collapsing at 100.0% and MCR fixed at 0.0%. SVM (RBF), Random Forest, and MLP also show tightly bound intervals centered near 99.2%, indicating highly reliable results. KNN and Logistic Regression

present slightly wider intervals but remain well within a high-performance range. Gaussian NB shows greater variance but remains consistent with its expected performance range. The bootstrap results confirm model robustness and reproducibility, with narrow confidence intervals and strong overlap among top models. This further validates the reliability of the stacking ensemble framework and its superior generalization on real CKD data.

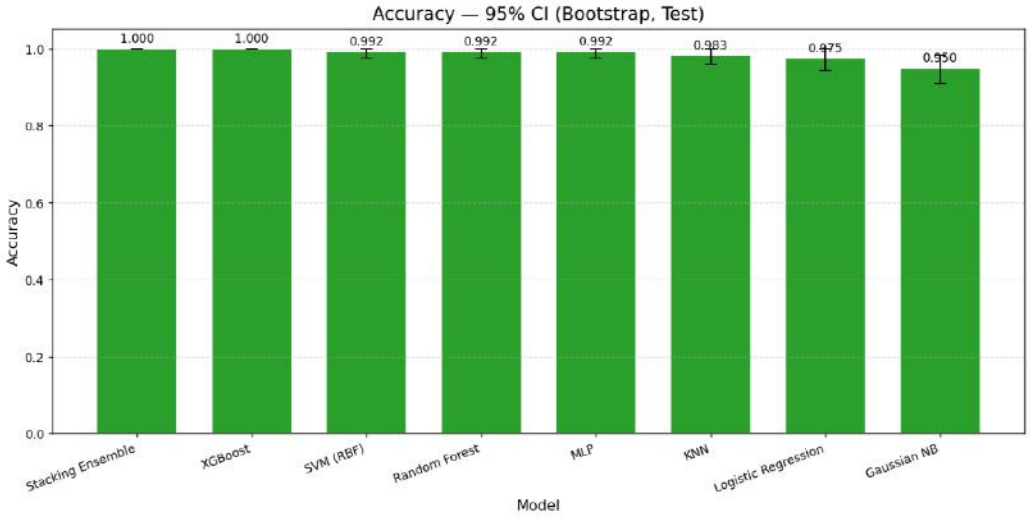

Figure 16: Bootstrap test on the UCI CKD dataset with 95% CI (Accuracy).

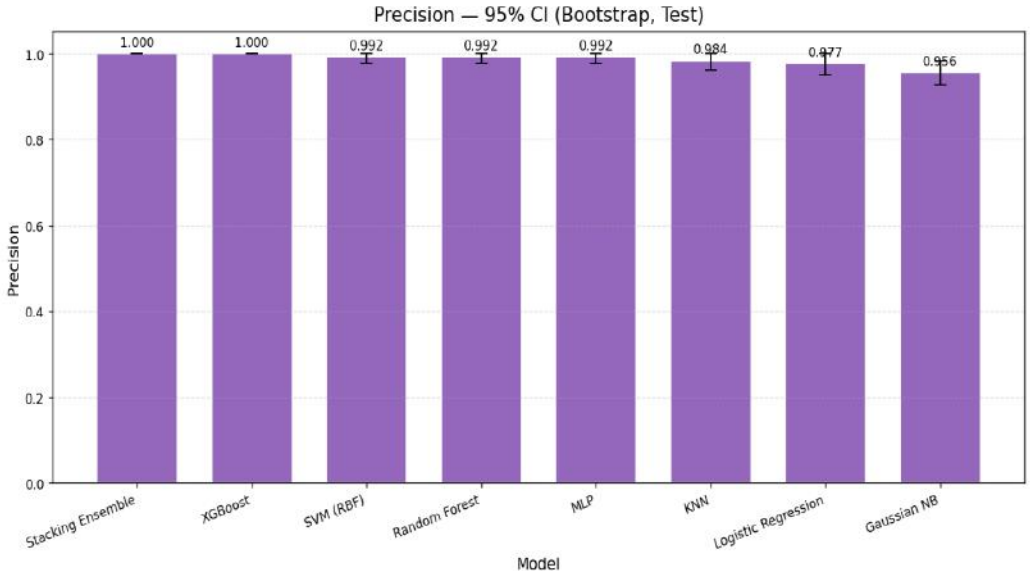

Figure 17: Bootstrap test on the UCI CKD dataset with 95% CI (Precision).

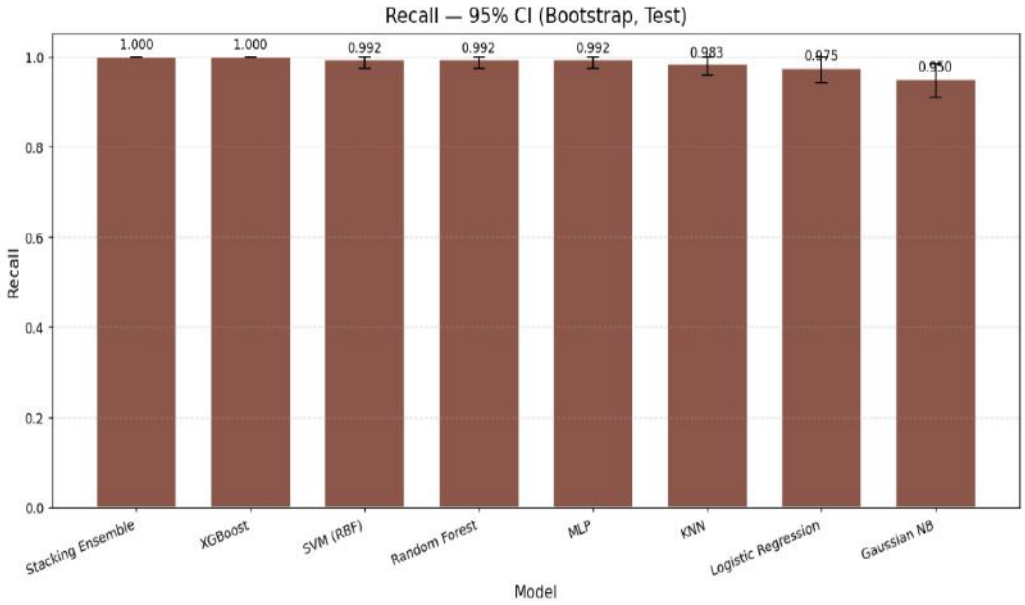

Figure 18: Bootstrap test on the UCI CKD dataset with 95% CI (Recall).

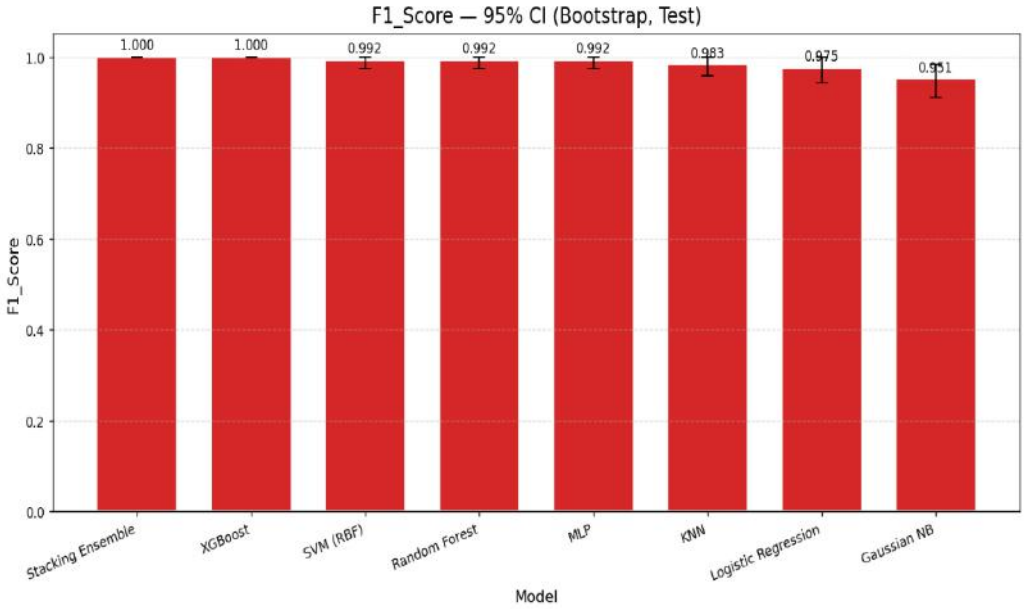

Figure 19: Bootstrap test on the UCI CKD dataset with 95% CI (F1_Score).

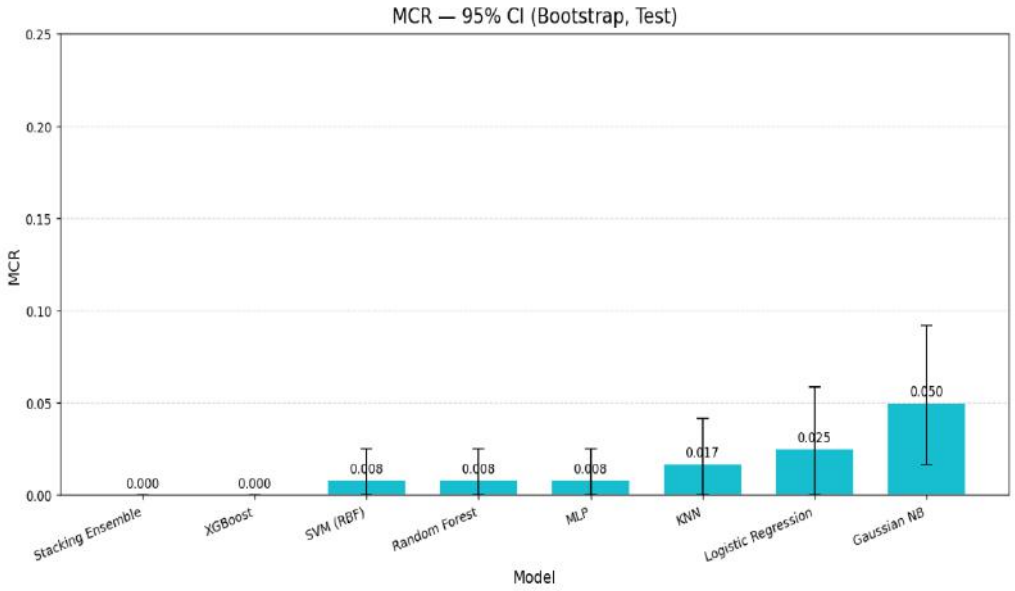

Figure 20: Bootstrap test on the UCI CKD dataset with 95% CI (MCR).

4.2 Performance on the Simulated CKD Dataset

To further validate the robustness and generalizability of the proposed framework, the models were evaluated on a simulated CKD dataset that mirrors real-world clinical variability while preserving the same feature schema as the UCI dataset. This experiment tests whether the identified predictive features and the stacking-based optimization remain effective under realistic noise, missingness, and correlation patterns.

4.2.1 Model Performance with Selected Features on the Simulated CKD Dataset

Table 10 summarizes test performance using the selected feature subset, while Figures 21–25 visualize the results across all evaluation metrics. The Stacking Ensemble, XGBoost, and MLP achieved the best overall accuracy of 96.7%, each with a misclassification rate (MCR) of 3.3%, confirming their resilience in synthetic, heterogeneous conditions. Logistic Regression followed closely at 96.3% (MCR = 3.7%), while Random Forest and SVM (RBF) performed comparably at 96.0% (MCR = 4.0%) and 95.7% (MCR = 4.3%), respectively. Gaussian NB maintained solid performance at 95.0%, and KNN, though lower at 92.0%, still demonstrated consistent classification. Precision, recall, and F1-score closely tracked accuracy across all models, indicating balanced prediction with minimal bias between CKD and non-CKD classes. Overall, the results show that the selected risk factors remain informative even under simulated variability, and the stacking ensemble sustains top-tier performance, reinforcing its stability beyond the constraints of a fixed clinical dataset. The narrow spread among leading models suggests that the feature selection strategy captured essential physiological indicators consistently, even when relationships among predictors were perturbed. This stability highlights the stacking ensemble's adaptability to complex, uncertain data environments, an important trait for future real-world deployment.

Table 10: Performance metrics for models using selected features on the simulated CKD dataset.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score %	MCR (%)
KNN	92.0%	92.4%	92.0%	92.0%	8.0%
SVM (RBF)	95.7%	95.7%	95.7%	95.7%	4.3%
Random Forest	96.0%	96.0%	96.0%	96.0%	4.0%
Gaussian NB	95.0%	95.2%	95.0%	95.0%	5.0%
Logistic Regression	96.3%	96.4%	96.3%	96.3%	3.7%
MLP	96.7%	96.7%	96.7%	96.7%	3.3%
XGBoost	96.7%	96.7%	96.7%	96.7%	3.3%
Stacking Ensemble	96.7%	96.7%	96.7%	96.7%	3.3%

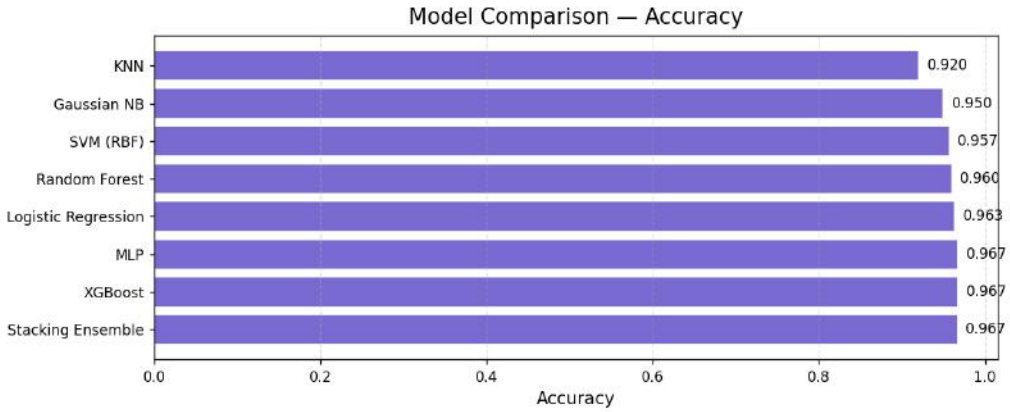

Figure 21: Comparative bar chart of model accuracy on the simulated CKD dataset.

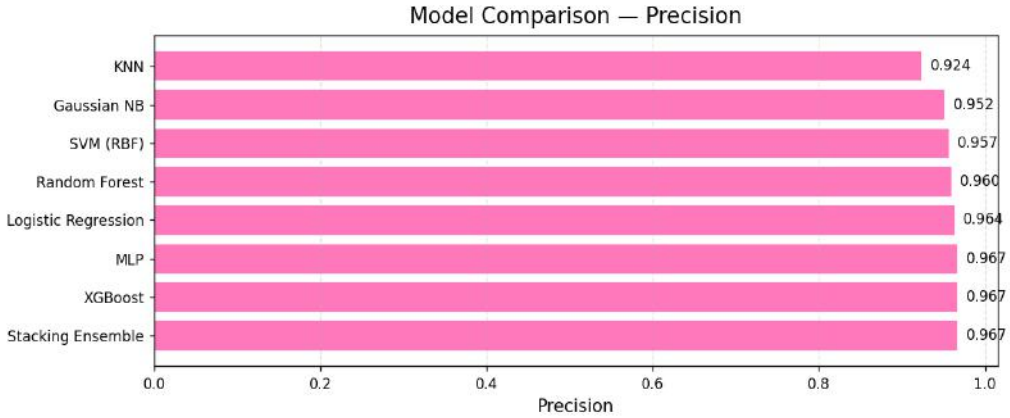

Figure 22: Comparative bar chart of model precision on the simulated CKD dataset.

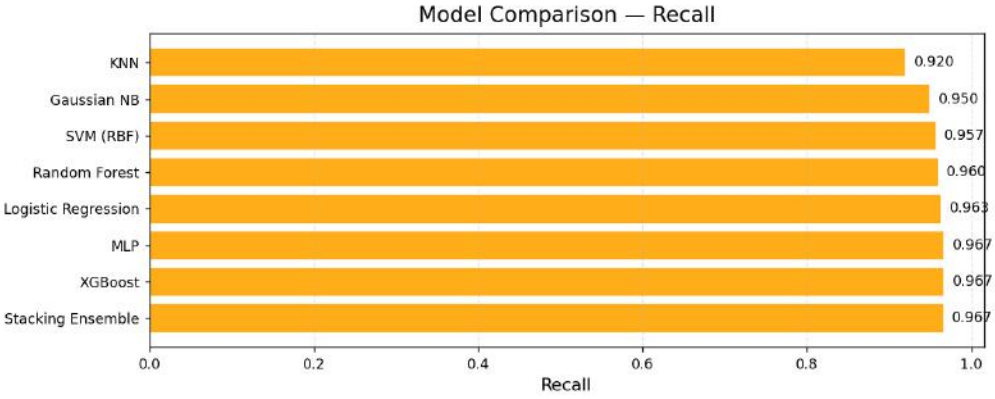

Figure 23: Comparative bar chart of model recall on the simulated CKD dataset.

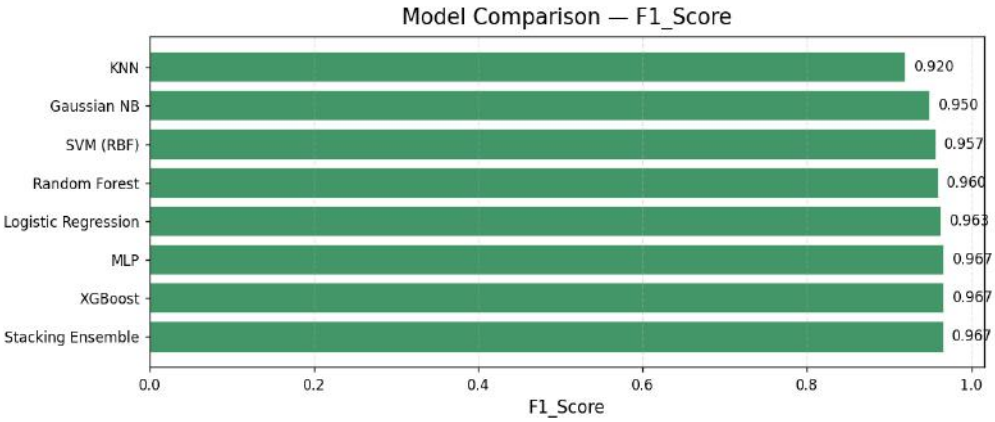

Figure 24: Comparative bar chart of model F1_Score on the simulated CKD dataset.

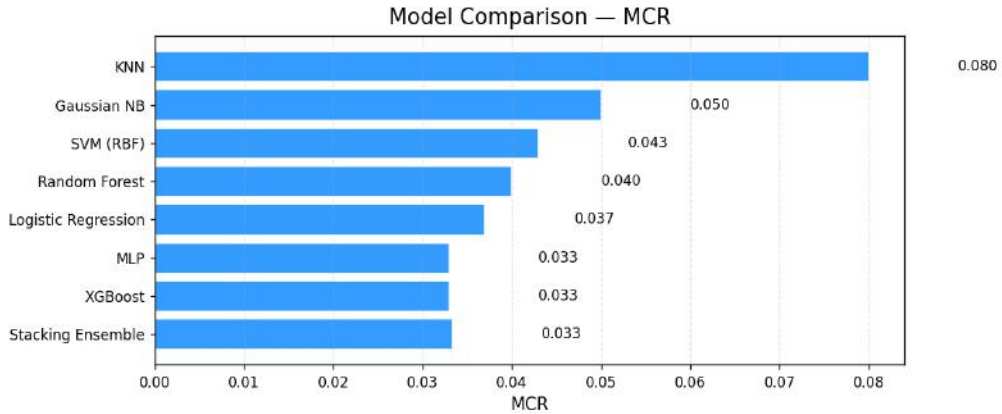

Figure 25: Comparative bar chart of model MCR on the simulated CKD dataset.

4.2.2 Bootstrap Performance with 95% Confidence Intervals on the Simulated CKD Dataset

Figures 26–30 display bootstrap-derived 95% confidence intervals (CIs) for all evaluation metrics. The Stacking Ensemble, XGBoost, and MLP remain statistically tied at the top, each centered at 96.7% accuracy with tight confidence intervals (94.7–98.7%) and MCR ≈ 3.3% (1.3–5.3%), indicating strong consistency and low variance. Logistic Regression, Random Forest, and SVM (RBF) follow closely, maintaining mean accuracies above 95% with slightly wider but overlapping CIs. Gaussian NB shows moderate dispersion (95.0%, MCR = 5.0%) and KNN exhibits the widest intervals (92.0%, MCR = 8.0%), reflecting sensitivity to feature scaling and local variance in the simulated data. Across all metrics, the CIs for precision, recall, and F1 closely follow those of accuracy, reinforcing that the models' predictive balance remains consistent under repeated resampling. These findings demonstrate that the proposed feature selection and stacking framework is both stable and generalizable, effectively extending predictive reliability beyond the empirical CKD data.

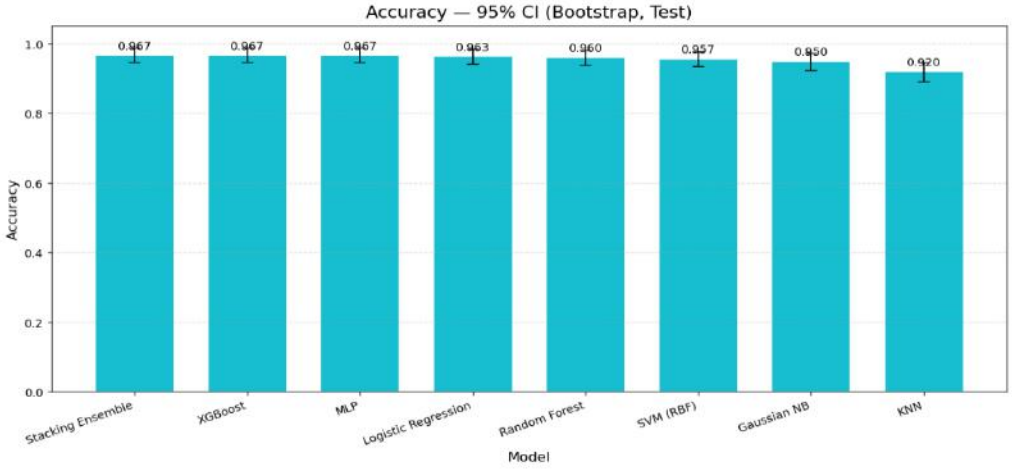

Figure 26: Bootstrap test on the simulated CKD dataset with 95% CI (Accuracy).

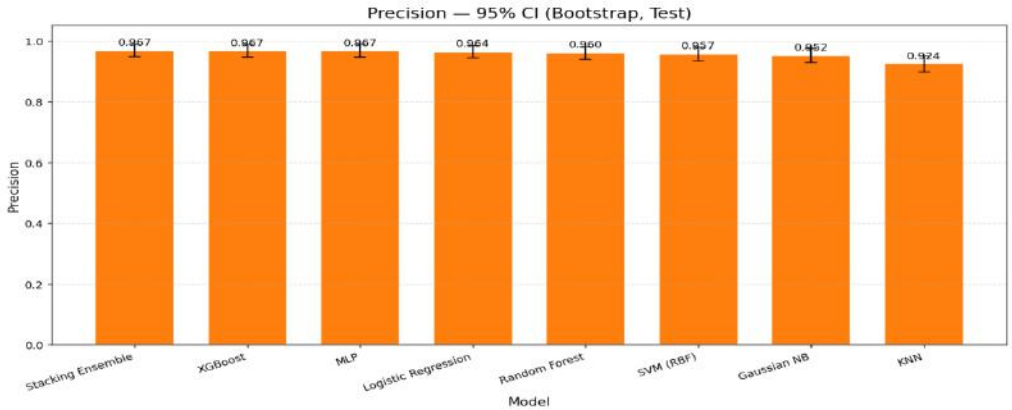

Figure 27: Bootstrap test on the simulated CKD dataset with 95% CI (Precision).

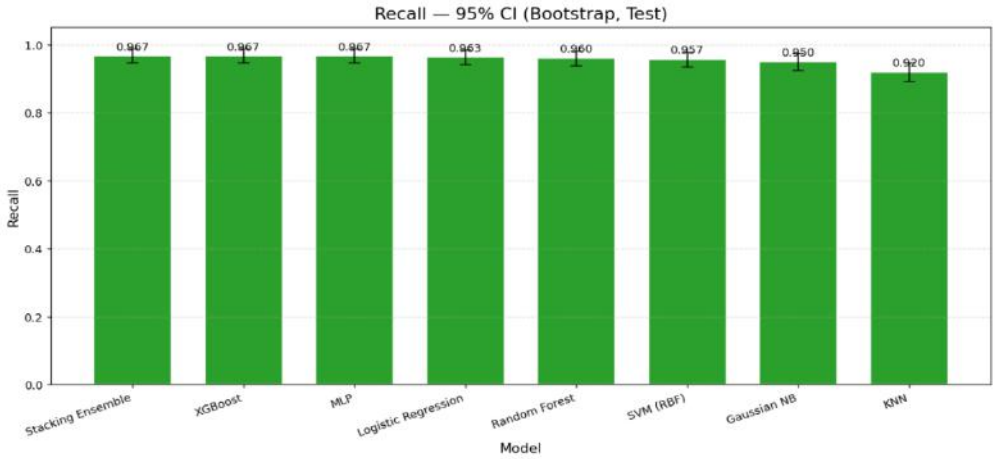

Figure 28: Bootstrap test on the simulated CKD dataset with 95% CI (Recall).

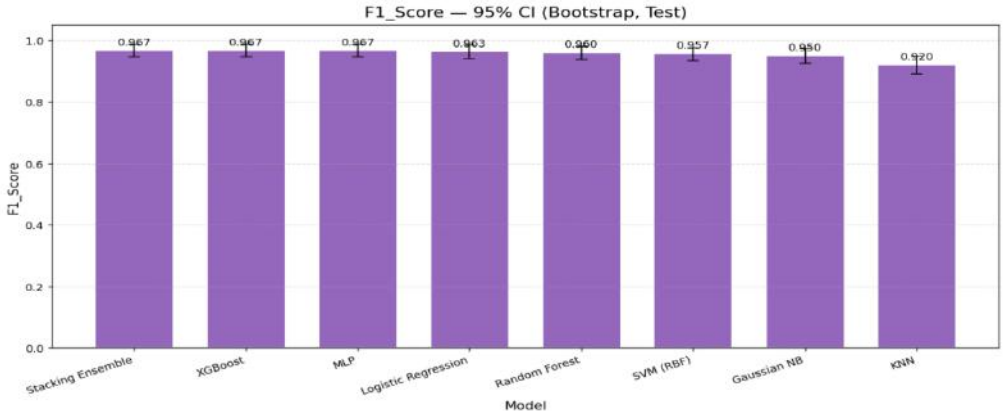

Figure 29: Bootstrap test on the simulated CKD dataset with 95% CI (F1_Score).

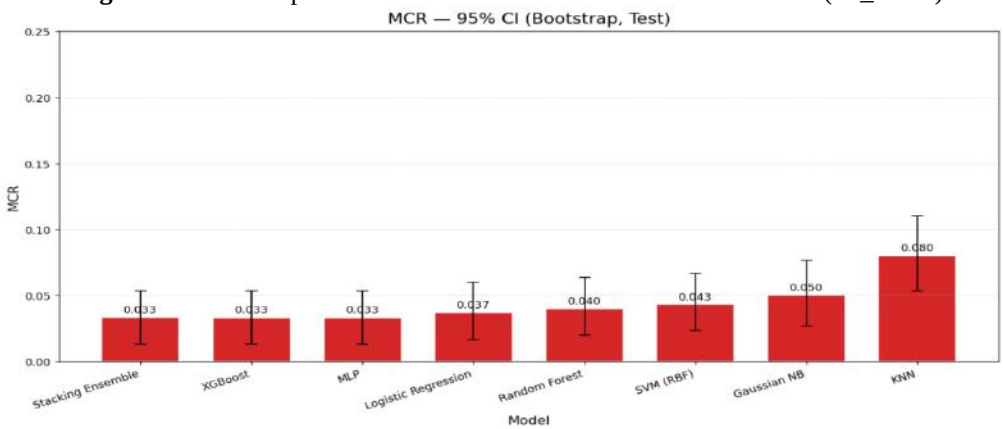

Figure 30: Bootstrap test on the simulated CKD dataset with 95% CI (MCR).

4.3 Accuracy Comparison between Datasets

Table 11 compares model accuracy across the UCI CKD and simulated CKD datasets to evaluate the stability and generalizability of the proposed framework. On the UCI dataset, all models achieved exceptionally high performance: Stacking and XGBoost reached 100.0% accuracy, MLP, Random Forest, and SVM (RBF) achieved 99.2%, Logistic Regression maintained 97.5%, KNN achieved 98.3%, and Gaussian NB followed with 95.0%. On the simulated dataset which introduced controlled noise, correlated variability, and mixed missingness patterns performance decreased slightly, as expected under more realistic data conditions. Stacking, XGBoost, and MLP remained the most stable, each achieving 96.7% accuracy with only minor decline. Logistic Regression followed closely at 96.3%, while Random Forest and SVM (RBF) maintained competitive results at 96.0% and 95.7%, respectively. Gaussian NB remained consistent at 95.0%, whereas KNN showed the largest drop to 92.0%, reflecting its sensitivity to shifts in data distribution and local density. These outcomes highlight that the stacking ensemble and feature selection strategy preserve strong predictive capability across both curated and simulated datasets. The ensemble's minimal accuracy decline demonstrates its robustness to data heterogeneity and confirms its suitability for real-world CKD prediction, where variability and missingness are inevitable.

Table 11: Accuracy comparison between UCI CKD dataset and simulated CKD dataset.

Model	UCI CKD Dataset Accuracy (%)	Simulated CKD Dataset Accuracy (%)
KNN	98.3%	92.0%
SVM (RBF)	99.2%	95.7%
Random Forest	99.2%	96.0%
Gaussian NB	95.0%	95.0%
Logistic Regression	97.5%	96.3%
MLP	99.2%	96.7%
XGBoost	100.0%	96.7%
Stacking Ensemble	100.0%	96.7%

The study demonstrates that using stacking machine learning models and selective feature sets significantly enhances the accuracy of CKD prediction.

4.3 Benchmarking of the Proposed Method

The proposed stacking ensemble method is benchmarked against existing studies on CKD prediction. Most studies focus on traditional models such as LR, RF, and J48, with accuracy ranging from 85.5% to 100%. Techniques like feature selection, ensemble methods, and hybrid models are commonly used to improve performance. For instance, feature selection (Poonia et al. [3], Almustafa [13]) and ensemble methods (Jongbo et al. [6], Ekanayake and Herath [30]) demonstrate improvements in prediction accuracy, often achieving near-perfect results. However, many of these models either lack generalization or face challenges like overfitting (Krishnamurthy et al. [5], Singh et al. [24]). While deep learning models show high accuracy, they are computationally expensive and may not always generalize well. Moreover, class imbalance handling through techniques like SMOTE (Nishat et al. [4], Ghosh et al. [18]) is not universally

applied, and hybrid models (Khalid et al. [12], M. Rashed-Al-Mahfuz et al. [23]) can be opaque in terms of model integration. The proposed stacking ensemble method addresses these gaps by leveraging multiple models, ensuring higher accuracy and better generalization, and providing a robust solution for CKD prediction across diverse datasets.

5. Conclusion

This study set out to do two things: pin down the predisposing risk factors that matter most for CKD, and use them to build a stronger predictor. We designed a leakage-safe preprocessing pipeline (train-only Z-score capping, median/mode imputation, min–max scaling, and SMOTE applied only on the training split), ran six complementary feature selection methods (LR, RFE, RF, MI, χ^2, PCA), and kept variables that appeared in a majority of methods. On the UCI CKD data, the common set was: serum creatinine, hemoglobin, packed cell volume, red blood cell count, specific gravity, albumin, sugar, hypertension, diabetes mellitus, and appetite. On the simulated CKD cohort, the set was: blood glucose random, blood urea, serum creatinine, potassium, hemoglobin, packed cell volume, specific gravity, albumin, red blood cells, pus cell clumps, appetite, pedal edema, and anemia. These selections align with renal physiology (filtration, concentration, anemia, and metabolic comorbidities), which supports their clinical face validity. Using these selected features, the stacking ensemble consistently delivered the best or tied-best results. On UCI CKD, stacking and XGBoost achieved 100.0% test accuracy, with SVM, Random Forest, and MLP close behind (≈99.2%), and minimal train–test gaps. On the simulated cohort, stacking, XGBoost, and MLP reached 96.7%, with Logistic Regression at 96.3%, and Random Forest/SVM near 96.0%/95.7%. Bootstrap 95% confidence intervals were tight for the leading models across both datasets, indicating stable estimates rather than favorable splits. In short, a compact, clinically interpretable feature set combined with stacking yields high accuracy and strong generalization on curated data and remains robust under realistic simulated variability. This balance of parsimony, transparency, and performance makes the approach well-suited for decision support. Results are based on one public dataset and a simulated cohort; external, multi-site validation and calibration analysis are needed. Future work will evaluate prospective drift, add explanation-at-prediction for clinicians, and test workflow integration. Careful preprocessing and majority-vote feature selection identify a small, clinically sensible set of CKD predictors. A stacking ensemble trained on those predictors offers state-of-the-art accuracy on UCI CKD and remains strong on realistic simulations an encouraging path toward reliable, usable CKD screening tools.

Authors' Contributions: Md. Razu Ahmed: Writing - original draft, formal analysis, data curation, visualization, conceptualization. Md. Abdur Rakib: Writing - review and editing, conceptualization, investigation. Abu Bakar Shiddik: Writing - review and editing, conceptualization, investigation. Md. Shamim Reza: Conceptualization, methodology, supervision, writing - review and editing.

Funding Details: This research did not receive funding from any specific grant agency, whether public, commercial, or non-profit.

Ethics Statement: This study utilizes a publicly available secondary dataset from the UCI ML repository. As such, ethical approval was not required for this dataset.

Disclosure statement: The authors declare no potential conflicts of interest related to this research.

Data Availability Statement: The data that support the findings of this study are openly available in UCI ML repository at https://archive.ics.uci.edu/dataset/336/chronic+kidney+disease

Acknowledgement: The author sincerely acknowledges the reviewer's valuable comments and constructive suggestions, which have greatly improved the quality of this manuscript.

References

[1] S. Roumeliotis, F. Mallamaci, and C. Zoccali (2020). Endothelial Dysfunction in Chronic Kidney Disease, from Biology to Clinical Outcomes: A 2020 Update, Journal of Clinical Medicine, vol. 9, no. 8, pp. 1–27, Jul. 2020, https://doi.org/10.3390/jcm9082359

[2] M. Murton et al., (2021). Burden of Chronic Kidney Disease by KDIGO Categories of Glomerular Filtration Rate and Albuminuria: A Systematic Review, Advances in Therapy, no. 38, pp. 180–200, Jan. 2021, https://doi.org/10.1007/s12325-020-01568-8

[3] R. C. Poonia et al., (2022). Intelligent Diagnostic Prediction and Classification Models for Detection of Kidney Disease, Healthcare, vol. 10, no. 2, p. 371, Feb. 2022, https://doi.org/10.3390/healthcare10020371

[4] M. M. Nishat et al., (2021). A Comprehensive Analysis on Detecting Chronic Kidney Disease by Employing Machine Learning Algorithms, EAI Endorsed Transactions on Pervasive Health and Technology, vol. 7, no. 29, pp. 1–12, Nov. 2021, https://doi.org/10.4108/eai.13-8-2021.170671

[5] S. Krishnamurthy et al., (2021). Machine Learning Prediction Models for Chronic Kidney Disease Using National Health Insurance Claim Data in Taiwan, Healthcare, vol. 9, no. 5, p. 546, May 2021, https://doi.org/10.3390/healthcare9050546

[6] O. A. Jongbo, A. O. Adetunmbi, R. B. Ogunrinde, and B. Badeji-Ajisafe (2020). Development of an ensemble approach to chronic kidney disease diagnosis, Scientific African, vol. 8, p. e00456, Jul. 2020, https://doi.org/10.1016/j.sciaf.2020.e00456

[7] T. Qin, L. Wu, Q. Hua, Z. Song, Y. Pan, and T. Liu (2020). Prediction of the mechanisms of action of Shenkang in chronic kidney disease: A network pharmacology study and experimental validation, Journal of Ethnopharmacology, vol. 246, p. 112128, Jan. 2020, https://doi.org/10.1016/j.jep.2019.112128

[8] H. Ilyas et al., (2021). Chronic kidney disease diagnosis using decision tree algorithms, BMC Nephrology, vol. 22, no. 1, p. 273, Dec. 2021, https://doi.org/10.1186/s12882-021-02474-z

[9] E. M. Senan et al., (2021). Diagnosis of Chronic Kidney Disease Using Effective Classification Algorithms and Recursive Feature Elimination Techniques, Journal of Healthcare Engineering, vol. 2021, no. 1, p. 1004767, Jun. 2021, https://doi.org/10.1155/2021/1004767

[10] B. Khan, R. Naseem, F. Muhammad, G. Abbas, and S. Kim (2020). An empirical evaluation of machine learning techniques for chronic kidney disease prophecy, IEEE Access, vol. 8, pp. 55012–55022, Mar. 2020, https://doi.org/10.1109/ACCESS.2020.2981689

[11] M. M. Nishat et al., (2020). Performance Investigation of Different Boosting Algorithms in Predicting Chronic Kidney Disease, IEEE, Dhaka, Dec. 2020. https://doi.org/10.1109/STI50764.2020.9350440

[12] H. Khalid, A. Khan, M. Zahid Khan, G. Mehmood, and M. Shuaib Qureshi (2023). Machine Learning Hybrid Model for the Prediction of Chronic Kidney Disease, Computational Intelligence and Neuroscience, vol. 2023, no. 1, p. 9266889, Mar. 2023, https://doi.org/10.1155/2023/9266889

[13] K. M. Almustafa (2021). Prediction of chronic kidney disease using different classification algorithms, Informatics in Medicine Unlocked, vol. 24, p. 100631, Jan. 2021, https://doi.org/10.1016/j.imu.2021.100631

[14] S. Dutta, R. Sikder, R. Islam, A. al Mukaddim, M. A. Hider, and M. Nasiruddin (2024). Comparing the Effectiveness of Machine Learning Algorithms in Early Chronic Kidney Disease Detection, Journal of Computer Science and Technology Studies, vol. 6, no. 4, pp. 77–91, Oct. 2024, https://doi.org/10.32996/jcsts.2024.6.4.11

[15] R. Gupta, N. Koli, N. Mahor, and N. Tejashri (2020). Performance Analysis of Machine Learning Classifier for Predicting Chronic Kidney Disease, IEEE, Belgaum, India, Jun. 2020. https://doi.org/10.1109/INCET49848.2020.9154147

[16] S. Yunus Yashfi, N. Sakib, S. Salman Pantho, M. Ashikul Islam, T. Islam, and M. Shahbaaz (2020). Risk Prediction of Chronic Kidney Disease Using Machine Learning Algorithms, IEEE, Kharagpur, India, Jul. 2020. https://doi.org/10.1109/ICCCNT49239.2020.9225548

[17] S. Umme Habiba et al., (2024). Early Prediction of Chronic Kidney Disease Using Machine Learning Algorithms with Feature Selection Techniques, in Communications in Computer and Information Science, Dubai: Springer Science and Business Media Deutschland GmbH, Aug. 2024, pp. 224–242. https://doi.org/10.1007/978-3-031-68639-9_14

[18] P. Ghosh, F. M. Javed Mehedi Shamrat, S. Shultana, S. Afrin, A. A. Anjum, and A. A. Khan (2020). Optimization of Prediction Method of Chronic Kidney Disease Using Machine Learning Algorithm, in Proceedings - 2020 15th International Joint Symposium on Artificial Intelligence and Natural Language Processing, iSAI-NLP 2020, Bangkok, Thailand: Institute of Electrical and Electronics Engineers Inc., Nov. 2020, pp. 1–6. https://doi.org/10.1109/iSAI-NLP51646.2020.9376787

[19] S. Akter et al., (2021). Comprehensive Performance Assessment of Deep Learning Models in Early Prediction and Risk Identification of Chronic Kidney Disease, IEEE Access, vol. 9, pp. 165184–165206, Nov. 2021, https://doi.org/10.1109/ACCESS.2021.3129491

[20] S. A. Ebiaredoh-Mienye, T. G. Swart, E. Esenogho, and I. D. Mienye (2022). A Machine Learning Method with Filter-Based Feature Selection for Improved Prediction of Chronic Kidney Disease, Bioengineering, vol. 9, no. 8, p. 350, Aug. 2022, https://doi.org/10.3390/bioengineering9080350

[21] B. P. Ghosh et al., (2024). Advancing Chronic Kidney Disease Prediction: Comparative Analysis of Machine Learning Algorithms and a Hybrid Model, Journal of Computer Science and Technology Studies, vol. 6, no. 3, pp. 15–21, Jun. 2024, https://doi.org/10.32996/jcsts.2024.6.3.2

[22] M. Hosseinzadeh et al., (2021). A diagnostic prediction model for chronic kidney disease in internet of things platform, Multimedia Tools and Applications, vol. 80, no. 11, pp. 16933–16950, May 2021, https://doi.org/10.1007/s11042-020-09049-4

[23] M. Rashed-Al-Mahfuz, A. Haque, A. Azad, S. A. Alyami, J. M. W. Quinn, and M. A. Moni (2021). Clinically Applicable Machine Learning Approaches to Identify Attributes of Chronic Kidney Disease (CKD) for Use in Low-Cost Diagnostic Screening, IEEE Journal of Translational Engineering in Health and Medicine, vol. 9, pp. 1–11, Apr. 2021, https://doi.org/10.1109/JTEHM.2021.3073629

[24] V. Singh, V. K. Asari, and R. Rajasekaran (2022). A Deep Neural Network for Early Detection and Prediction of Chronic Kidney Disease, Diagnostics, vol. 12, no. 1, p. 116, Jan. 2022, https://doi.org/10.3390/diagnostics12010116

[25] P. Chittora et al., (2021). Prediction of Chronic Kidney Disease - A Machine Learning Perspective, IEEE Access, vol. 9, pp. 17312–17334, Jan. 2021, https://doi.org/10.1109/ACCESS.2021.3053763

[26] M. A. Islam, S. Akter, M. S. Hossen, S. A. Keya, S. A. Tisha, and S. Hossain (2020). Risk factor prediction of chronic kidney disease based on machine learning algorithms, in Proceedings of the 3rd International Conference on Intelligent Sustainable Systems, ICISS 2020, Thoothukudi, India: Institute of Electrical and Electronics Engineers Inc., Dec. 2020, pp. 952–957. https://doi.org/10.1109/ICISS49785.2020.9315878

[27] E. Dritsas and M. Trigka (2022). Machine Learning Techniques for Chronic Kidney Disease Risk Prediction, Big Data and Cognitive Computing, vol. 6, no. 3, p. 98, Sep. 2022, https://doi.org/10.3390/bdcc6030098

[28] M. A. Islam, M. Z. H. Majumder, and M. A. Hussein (2023). Chronic kidney disease prediction based on machine learning algorithms, Journal of Pathology Informatics, vol. 14, p. 100189, Jan. 2023, https://doi.org/10.1016/j.jpi.2023.100189

[29] V. Srikanth (2023). Chronic Kidney Disease Prediction Using Machinelearning Algorithms, International Journal of Techno-Engineering, vol. XV, no. I, Jan. 2023.http://ijte.uk/archive/2023/CHRONIC-KIDNEY-DISEASE-PREDICTION-USING-MACHINE-LEARNING-ALGORITHMS.pdf

[30] I. Udara Ekanayake and D. Herath (2020). Chronic Kidney Disease Prediction Using Machine Learning Methods, IEEE, Moratuwa, Sri Lanka, Jul. 2020. https://doi.org/10.1109/MERCon50084.2020.9185249

[31] B. Gudeti, S. Mishra, S. Malik, T. F. Fernandez, A. K. Tyagi, and S. Kumari (2020). A Novel Approach to Predict Chronic Kidney Disease using Machine Learning Algorithms, in Proceedings of the 4th International Conference on Electronics, Communication and Aerospace Technology, ICECA 2020, Coimbatore, India: Institute of Electrical and Electronics Engineers Inc., Nov. 2020, pp. 1630–1635. https://doi.org/10.1109/ICECA49313.2020.9297392

[32] R. Sawhney, A. Malik, S. Sharma, and V. Narayan (2023). A comparative assessment of artificial intelligence models used for early prediction and evaluation of chronic kidney disease, Decision Analytics Journal, vol. 6, p. 100169, Mar. 2023, https://doi.org/10.1016/j.dajour.2023.100169

[33] Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group (2024). KDIGO 2024 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease, Kidney International, vol. 105, no. 4S, pp. S117–S314, Apr. 2024, https://doi.org/10.1016/j.kint.2023.10.018

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