Discover Artificial Intelligence

RESEARCH Open Access

A privacy preserving federated meta ensemble stacking framework with explainability and differential privacy for chronic kidney disease prediction in resource constrained settings

Komal Kumar Napa[1], D. Sathyanarayanan[2], Raguraman Purushothaman[3], Haiter Lenin Allasi[4*], Mary Vasanthi Soosaimariyan[5], Cornelius Karunakaran[6] and M. Somasundara Rao[7]

*Correspondence:
Haiter Lenin Allasi
drahlenin@kiot.edu.et;
haiterlenina@gmail.com

Full list of author information is available at the end of the article

Abstract

Chronic kidney disease (CKD) is a major public health concern that requires early and reliable diagnosis to reduce disease progression and associated complications. Existing machine learning approaches often rely on centralized training, limiting their applicability in healthcare environments where data privacy, heterogeneity, and resource constraints are critical concerns. To address these challenges, this study proposes a privacy-preserving federated meta-ensemble stacking framework for CKD prediction. The proposed approach integrates a hybrid preprocessing pipeline, federated learning with prediction-level aggregation, differential privacy, and SHAP-based explainability to enable secure, interpretable, and decentralized model learning. Experimental evaluation demonstrates a mean cross-validation accuracy of 98.85% with an F1-score of 98.58%. The framework maintains robust performance under Gaussian noise, achieving 97.50% accuracy, while cross-client evaluation reaches 99.25%, demonstrating strong generalization across distributed healthcare institutions. The model also exhibits excellent discrimination and calibration, achieving a ROC-AUC of 0.998, PR-AUC of 0.997, and a Brier score of 0.00038. Furthermore, communication overhead is reduced from 0.82 MB to 0.25 MB through compression and sparse updates. These findings demonstrate that the proposed framework provides an accurate, privacy-aware, and interpretable solution for decentralized CKD prediction.

Keywords Chronic kidney disease, Federated learning, Differential privacy, Machine learning, Explainable artificial intelligence, Resource-constrained environments

© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

1 Introduction

Chronic kidney disease (CKD) is a progressive and often irreversible condition that affects millions of individuals worldwide and places a significant burden on healthcare systems. Early detection is critical, as timely intervention can slow disease progression and reduce the risk of severe complications such as kidney failure and cardiovascular disorders. However, CKD is frequently under diagnosed due to subtle early symptoms and reliance on manual clinical assessments, which may vary across practitioners [1]. In recent years, machine learning (ML) techniques have shown considerable promise in assisting clinical decision-making by enabling automated, data-driven prediction of CKD using patient records and laboratory measurements [2].

Despite these advances, most existing CKD prediction models are developed in centralized settings, where data from multiple sources are aggregated into a single repository for training [3]. While this approach often yields high predictive performance, it raises serious concerns regarding data privacy, security, and regulatory compliance, particularly in the healthcare domain where sensitive patient information is involved [4]. Moreover, centralized models do not reflect real-world scenarios, as hospitals and medical institutions are often unable or unwilling to share raw data due to legal and ethical constraints [5]. As a result, there is a growing need for decentralized learning paradigms that can enable collaborative model development without compromising data privacy.

Federated learning (FL) has emerged as a promising solution to address these challenges by allowing multiple institutions to collaboratively train machine learning models while keeping sensitive data locally at each client [6]. Recent studies have further demonstrated the effectiveness of federated learning for secure distributed intelligence by combining privacy-preserving model aggregation with robust learning strategies capable of maintaining high predictive performance under heterogeneous client configurations and imbalanced data distributions [7]. In parallel, recent advances in intelligent security systems have shown that integrating data balancing, feature reduction, and ensemble learning significantly improves robustness, scalability, and detection performance in distributed environments [8, 9]. Despite these developments, federated learning systems continue to face challenges related to heterogeneous data distributions, communication overhead, privacy leakage through shared model information, and deployment in resource-constrained environments [10]. Furthermore, many existing FL-based healthcare studies focus primarily on predictive performance while paying limited attention to model interpretability, robustness under noisy conditions, and communication efficiency. Explainability is particularly important in healthcare because clinicians require transparent and clinically meaningful explanations before adopting AI-assisted decision support systems [11].

Another limitation of current approaches lies in their reliance on single models or simplistic ensembles, which may not fully capture complex patterns in clinical data. Additionally, although privacy-preserving techniques such as differential privacy have been investigated, they are rarely integrated with advanced ensemble learning and explainability within a unified federated framework [12]. Consequently, there remains a clear research gap in developing an end-to-end framework that simultaneously addresses predictive performance, privacy preservation, interpretability, robustness, and communication efficiency for decentralized CKD prediction [13].

The key contributions of this work can be summarized as follows:

1. A prediction-level federated meta-ensemble stacking framework for CKD prediction that preserves data privacy by aggregating locally generated prediction probabilities instead of sharing raw patient data or model parameters.
2. Integration of differential privacy with temperature scaling to achieve a balance between data protection and predictive performance.
3. Implementation of federated explainability using SHAP to ensure consistent and interpretable feature importance across clients.
4. Comprehensive evaluation under realistic conditions, including cross-client validation, noise robustness, and privacy-constrained scenarios.
5. Analysis of communication efficiency to demonstrate the feasibility of deploying the framework in resource-constrained environments.
6. The proposed framework integrates diversity-aware preprocessing, federated meta-ensemble stacking, differential privacy, SHAP-based explainability, and communication optimization into a unified architecture for secure and interpretable CKD prediction.

1.1 Literature review

The application of machine learning techniques for chronic kidney disease (CKD) prediction has gained considerable attention in recent years due to their ability to uncover complex patterns in clinical data. Early studies primarily relied on traditional classification algorithms such as logistic regression, decision trees, and support vector machines, which demonstrated promising results in identifying CKD risk factors and improving diagnostic accuracy [14]. These approaches often emphasized feature selection and preprocessing techniques to enhance model performance; however, their effectiveness was sometimes limited by the complexity and variability of clinical datasets [15]. With the advancement of computational methods, ensemble learning techniques such as Random Forest and Gradient Boosting have been widely adopted, offering improved predictive capabilities by combining multiple learners [16].

More recently, deep learning models have been explored for CKD prediction, leveraging their ability to model nonlinear relationships and high-dimensional data [17]. Convolutional neural networks and hybrid architectures have been applied to medical imaging and tabular datasets, showing improved accuracy compared to traditional models [18]. Despite these advancements, deep learning approaches often require large volumes of data and computational resources, making them less suitable for smaller healthcare datasets or resource-limited settings [19]. Furthermore, many of these models operate as "black boxes," providing limited interpretability, which restricts their acceptance in clinical practice [20].

To overcome issues related to limited data availability and uneven data distribution, recent research has increasingly focused on the use of federated learning (FL) in healthcare domains. This approach allows multiple institutions to collaboratively develop predictive models without exchanging sensitive patient information, thereby supporting privacy preservation and compliance with regulatory requirements [21]. Several studies have highlighted the potential of FL in medical prediction tasks, including cardiovascular and diabetes-related applications, demonstrating its suitability for decentralized healthcare systems [22].

Despite these advantages, federated learning presents several technical challenges. Variations in data distributions across participating clients can lead to inconsistencies in model performance, while communication between clients and the central server may introduce additional overhead. Furthermore, achieving stable convergence becomes difficult when client data are non-identically distributed [23]. In addition, many existing FL-based models rely on basic aggregation techniques and often do not leverage more sophisticated ensemble learning strategies that could further enhance predictive accuracy and robustness.

In addition, methods such as differential privacy (DP) are used in machine learning to maintain data confidentiality. DP mechanisms introduce controlled noise into data or model parameters to prevent the leakage of individual-level information [24]. While DP enhances data security, it may also lead to a reduction in model accuracy if not carefully calibrated [25]. Existing studies often treat FL and DP as separate components, with limited integration into a unified framework that balances performance, interpretability and privacy.

Explainable artificial intelligence (XAI) has also emerged as a critical requirement in healthcare applications, where model transparency is essential for clinical trust and decision-making. Techniques such as SHAP and LIME have been widely used to interpret model predictions by quantifying feature contributions [26]. Although these methods provide valuable insights, most implementations are limited to centralized models, and there is a lack of research on explainability in federated settings [27]. This limitation restricts the ability to derive consistent and globally interpretable insights across distributed data sources.

Additionally, practical deployment considerations such as communication efficiency, computational complexity, and resource constraints are often overlooked in existing studies. Although recent research has demonstrated that hybrid learning models, advanced optimization techniques, and secure federated frameworks can substantially improve predictive performance and privacy preservation, most approaches continue to focus primarily on model accuracy rather than deployment feasibility in distributed healthcare environments [28–32]. Similarly, recent explainable and hybrid AI frameworks developed for disease diagnosis have highlighted the importance of feature selection, ensemble learning, and model interpretability; however, these methods are predominantly designed for centralized settings and do not address collaborative learning under strict privacy constraints [33–35]. Furthermore, federated systems may incur considerable communication overhead due to repeated exchanges of model information, making them less suitable for bandwidth-limited healthcare infrastructures [36]. Only a limited number of studies have investigated communication-efficient mechanisms such as model compression or sparse communication, leaving a significant gap in the development of scalable, privacy-preserving, and resource-efficient federated healthcare frameworks.

Although substantial progress has been achieved in CKD prediction using machine learning and deep learning techniques, existing approaches continue to exhibit limitations in privacy preservation, interpretability, robustness, communication efficiency, and real-world deployment. Moreover, very few studies integrate prediction-level federated aggregation, differential privacy, meta-ensemble learning, explainable AI, and communication-efficient strategies within a unified framework. These limitations motivate the

development of the proposed framework, which combines privacy-preserving collaborative learning, meta-ensemble stacking, federated SHAP explainability, and communication-efficient prediction aggregation to provide an accurate, interpretable, and scalable solution for CKD prediction.

Additionally, practical deployment considerations such as communication efficiency and resource constraints are often overlooked in existing studies. Federated systems can incur significant communication costs due to repeated model updates, making them less feasible for real-world applications in low-resource environments [36]. Few works address this issue through model compression or sparse communication strategies, leaving a gap in designing scalable and efficient federated healthcare solutions.

In summary, while substantial progress has been made in CKD prediction using machine learning, existing approaches face limitations in terms of privacy preservation, interpretability, robustness, and deployment feasibility. These gaps indicate the need for a combined framework that includes federated learning, privacy mechanisms, ensemble models, and explainability, while also handling practical challenges such as communication and limited resources.

2 Materials and methods

2.1 Overview of the proposed framework

This study presents a privacy-preserving federated meta-ensemble stacking framework for chronic kidney disease (CKD) prediction, designed to provide accurate, secure, and interpretable clinical decision support in decentralized healthcare environments. The overall workflow of the proposed framework is illustrated in Fig. 1, which depicts the complete pipeline from data preprocessing to federated model training, privacy preservation, explainability, and communication optimization. Initially, the raw CKD dataset undergoes a diversity-aware preprocessing stage involving hybrid imputation, categorical encoding, feature scaling, and mutual information-based feature selection to generate a clean and informative feature space. The processed data are subsequently partitioned across multiple clients, where an identical meta-ensemble stacking model

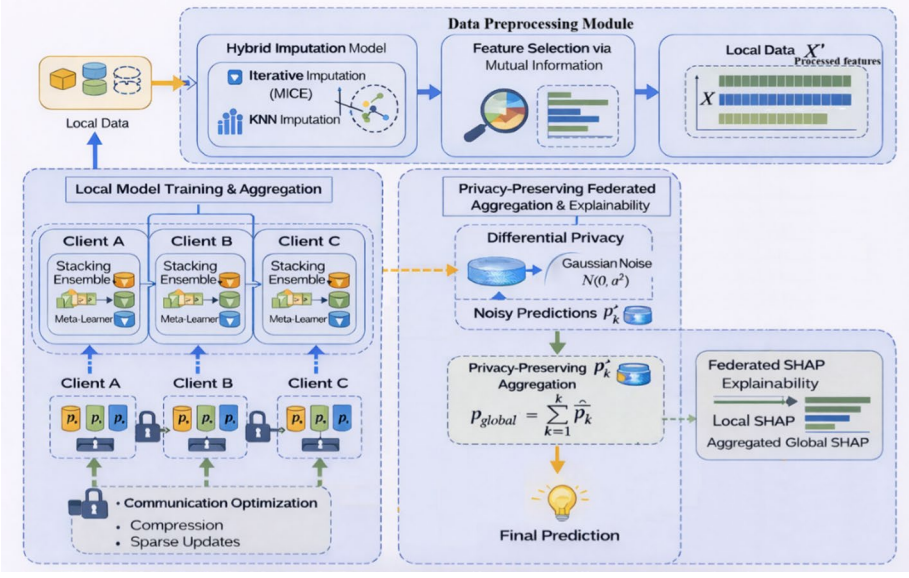


Fig. 1 Workflow of the proposed system

comprising Extra Trees, LightGBM, and CatBoost is independently trained at each institution. Instead of exchanging raw patient records, clients share only prediction probabilities, which are protected using differential privacy before being transmitted to the central server for prediction-level aggregation. The aggregated predictions are then used to generate the final CKD classification, while federated SHAP explainability provides both local and global interpretations of model decisions. To facilitate deployment in resource-constrained healthcare environments, communication overhead is further reduced through model compression and sparse updates. The complete procedure of the proposed framework is presented in Algorithm 1. The objective is to learn a predictive mapping:

$$f : X \rightarrow Y \tag{1}$$

where $X \in \mathbb{R}^{n \times d}$ represents the feature matrix and $Y \in \{0,1\}$ denotes CKD classification.

Algorithm 1 Proposed fed erated meta-ensemble framework

Input: CKD Dataset D
Output: CKD Prediction ŷ

1: Preprocess dataset
2: Impute missing values
3: Encode categorical variables
4: Scale numerical features
5: Select important features using MI
6: Partition dataset into K clients
7: for each client do
8: Train stacking model
9: Generate prediction probabilities
10: Apply Differential Privacy
11: end for
12: Aggregate client predictions
13: Compute final prediction
14: Compute Federated SHAP explanations
15: Apply communication optimization
16: Return prediction ŷ

2.2 Dataset description

The dataset [37] used in this study comprises 400 patient records with 26 clinical attributes representing physiological and laboratory measurements associated with chronic kidney disease (CKD). Although the dataset is relatively small, synthetic data augmentation was not employed to preserve the original clinical data distribution. The feature space can be formally defined as:

$$X \in \mathbb{R}^{400 \times 26} \tag{2}$$

where each row corresponds to an individual patient and each column represents a clinical variable. The target variable $Y \in \{0,1\}$ denotes CKD and non-CKD classes. Table 1 depicts the summary of the CKD dataset.

The dataset includes a mix of numerical (12 features) and categorical (14 features) attributes. Numerical features include age, blood pressure (bp), blood glucose random (bgr), serum creatinine (sc), sodium (sod), potassium (pot), and haemoglobin (hemo),

Table 1 Summary of the CKD dataset

Category	Description
Dataset size	400 samples
Number of features	26 (including target)
Input features	25
Target variable	Classification (CKD / Not CKD)
Numerical features	12
Categorical features	14
Data type	Mixed (Numerical + Categorical)
Missing values	Present (up to 38%)
Highest missing feature	RBC (38%)
Moderate missing features	RC (32.5%), WC (26.25%), Potassium (22%)
Low missing features	Age (2.25%), BP (3%), SC (4.25%)
Class distribution	CKD: 250, Not CKD: 150
Feature characteristics	Skewed distributions, correlated clinical variables
Data challenges	Missing values, class imbalance, mixed data types
Preprocessing required	Imputation, encoding, scaling, feature selection

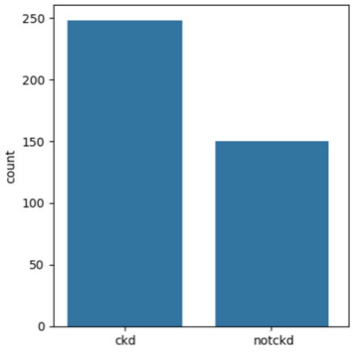


Fig. 2 Class distribution of the CKD dataset

among others. Categorical features capture clinical conditions such as hypertension (htn), diabetes mellitus (dm), and anemia (ane).

The class distribution (Fig. 2) indicates a moderate imbalance, with approximately 250 CKD cases and 150 non-CKD cases, reflecting realistic clinical prevalence patterns.

Exploratory analysis reveals several important characteristics. The correlation matrix (Fig. 3) indicates meaningful relationships among clinical variables, particularly between biochemical parameters such as serum creatinine, blood urea, and haemoglobin, which are known indicators of kidney function. Additionally, numerical feature distributions (Fig. 4) show varying levels of skewness, especially in attributes such as blood urea and serum creatinine, suggesting the need for transformation and scaling.

A key challenge in this dataset is the presence of missing values (Fig. 5). Several features exhibit substantial missingness, including red blood cells (rbc) with 38%, red cell count (rc) with 32.5%, and white blood cell count (wc) with 26.25%. This highlights the necessity for a robust preprocessing strategy to ensure reliable model performance.

2.3 Preprocessing

The CKD dataset exhibits considerable diversity in both feature characteristics and data quality, presenting several challenges commonly encountered in real-world clinical datasets. It comprises a heterogeneous combination of continuous laboratory measurements,

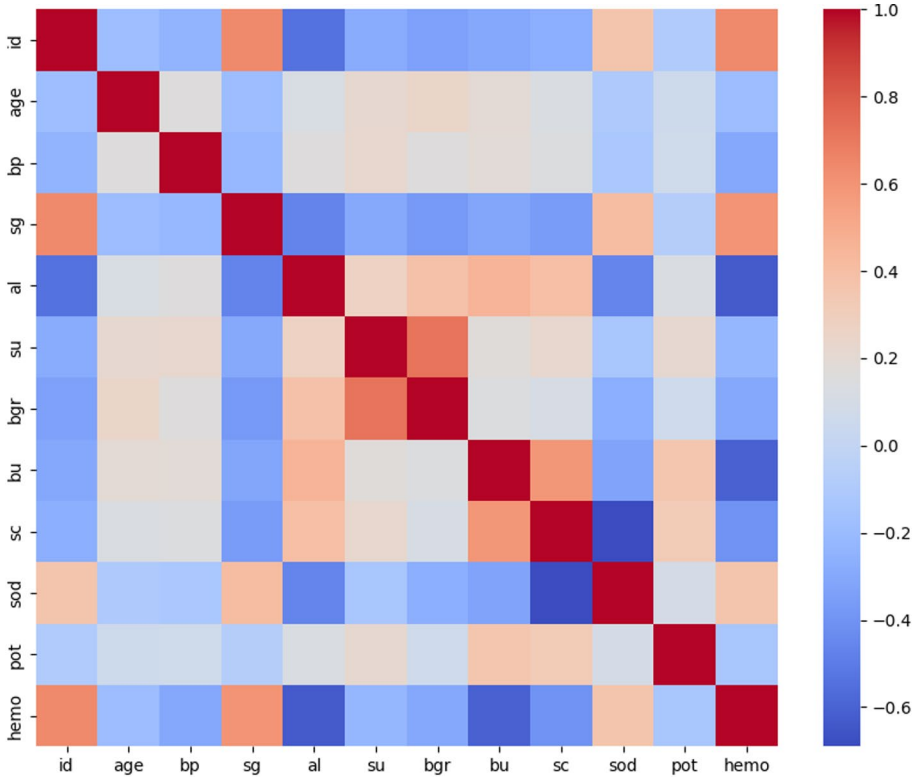


Fig. 3 Correlation matrix

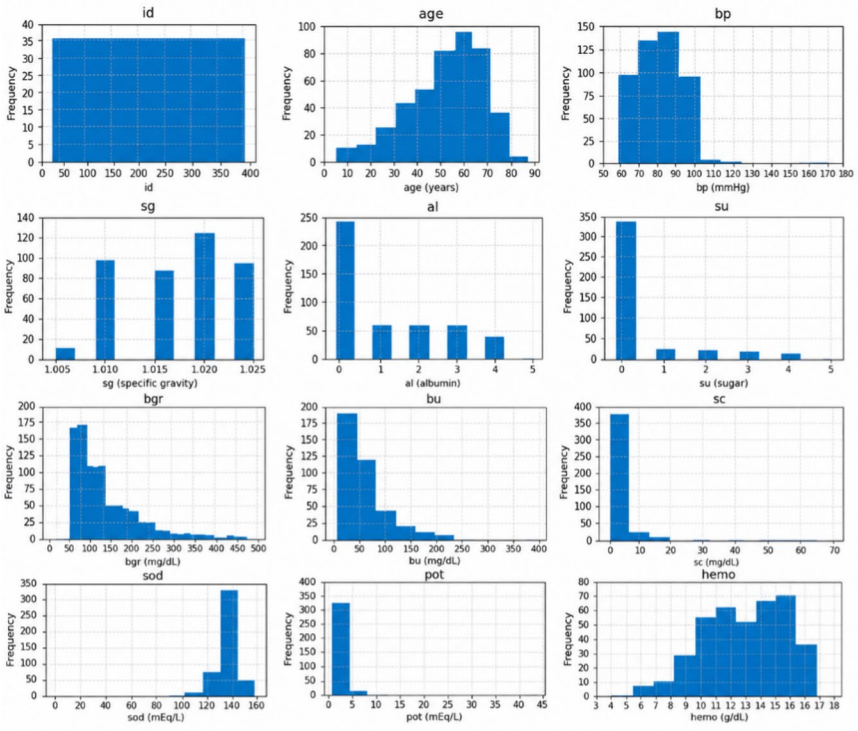


Fig. 4 Numerical feature distribution of the CKD dataset

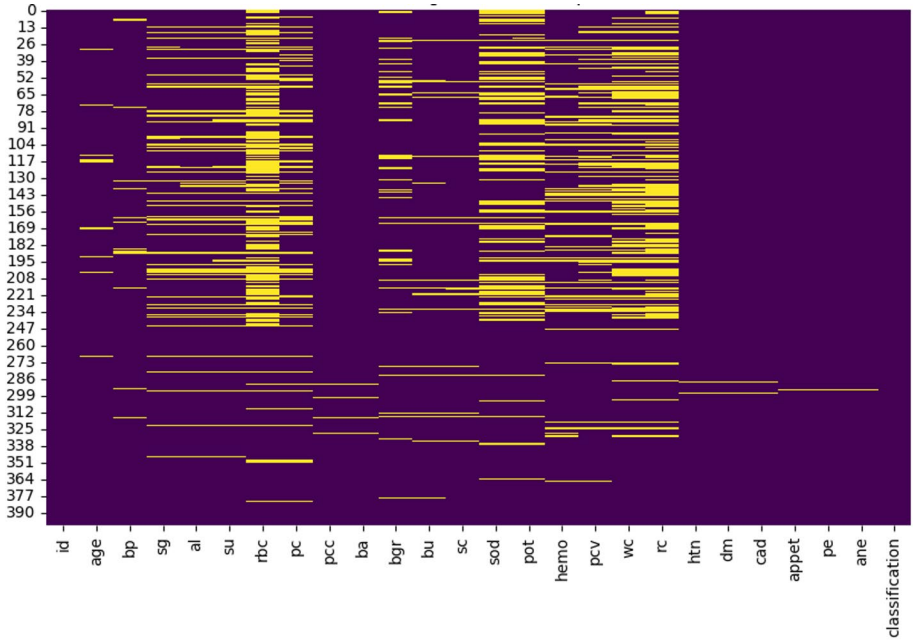


Fig. 5 Missing value heatmap

physiological variables, and categorical clinical indicators, each with distinct statistical properties. Moreover, the dataset contains varying levels of missing values across attributes, ranging from less than 1% in some variables to approximately 38% in others, along with skewed feature distributions, inconsistent categorical representations, and redundant information. Such heterogeneity can adversely affect model learning if a uniform preprocessing strategy is applied. Therefore, a structured multi-stage preprocessing pipeline was designed to address the specific characteristics of different feature groups while preserving clinically meaningful information.

Rather than employing a single preprocessing technique, the proposed pipeline adopts a diversity-aware strategy in which each operation is selected according to the underlying data characteristics. Numerical attributes with moderate missing values are reconstructed using iterative imputation to preserve global relationships among clinical variables, whereas features with substantial missingness are imputed using the K-nearest neighbors (KNN) algorithm to maintain local similarity between patient records. Categorical variables are transformed into numerical representations through label encoding, while highly skewed numerical features are normalized using robust scaling to reduce the influence of outliers and improve numerical stability. Finally, mutual information-based feature selection is applied to retain only the most informative clinical attributes associated with CKD prediction. This adaptive preprocessing framework effectively addresses the diverse characteristics of the dataset, resulting in a clean, consistent, and informative feature space that provides a reliable foundation for the subsequent federated meta-ensemble learning framework.

Accordingly, the preprocessing workflow was organized into a sequence of complementary stages, each designed to address a specific characteristic of the dataset while preserving clinically relevant information. The individual preprocessing steps are described in the following subsections.

2.3.1 Handling missing values

A major issue in the dataset is the presence of missing values across multiple features, with certain attributes such as *rbc*, *rc*, and *wc* exhibiting high missingness (38%, 32.5%, and 26.25%, respectively). Such incomplete data can significantly affect model performance if not handled appropriately. To address this issue, a hybrid imputation strategy is adopted. For features with moderate missing values, iterative imputation is applied, where each feature is estimated using the remaining observed features in the dataset. This process can be expressed as:

$$x_j = f_j\left(X_{\setminus j}\right) + \epsilon \tag{3}$$

where x_j denotes the feature being imputed, $X_{\setminus j}$ represents the set of observed features excluding x_j, $f_j(\cdot)$ is the predictive function used for iterative imputation, and ϵ denotes the residual error. For features with higher missing rates, K-nearest neighbors (KNN) imputation is employed, where the missing value of a sample is estimated as the average of its k nearest neighboring samples:

$$\hat{x}_i = \frac{1}{k} \sum_{j \in N_k(i)} x_j \tag{4}$$

where $\widehat{x}_i$ denotes the imputed value of the missing feature for sample i, $N_k(i)$ represents the set of the k nearest neighbors of sample i, x_j denotes the corresponding observed feature values of the neighboring samples, and k is the number of nearest neighbors. This approach preserves local similarity structures within the dataset by estimating missing values from clinically similar patient records.

As a result of this hybrid imputation strategy, all missing values were successfully eliminated, resulting in a NaN count of zero before feature selection while preserving both global feature dependencies and local neighborhood relationships.

2.3.2 Categorical encoding

In addition to missing data, the dataset includes several categorical variables such as *rbc*, *pc*, *htn*, and *dm*, which cannot be directly processed by machine learning algorithms. These variables are converted into numerical form using label encoding, defined as:

$$x_{cat} \rightarrow \{0, 1, 2, \dots\} \tag{5}$$

Binary categories (e.g., yes/no) are mapped into integer representations, ensuring consistency across the dataset. This transformation enables seamless integration with tree-based and ensemble learning models without any loss of information, while preserving the inherent structure of categorical relationships.

2.3.3 Data cleaning and consistency

The dataset also contains inconsistencies such as duplicate class labels (e.g., "ckd\t"), formatting issues, and mixed textual representations. These irregularities can introduce bias and reduce model reliability if left uncorrected. To resolve this, a normalization process is applied:

$$x_{clean} = \text{normalize}\left(x_{raw}\right) \tag{6}$$

This step involves standardizing categorical values, removing unwanted characters, and ensuring uniform labelling across all samples. Following this correction, the target distribution is refined to 250 CKD cases and 150 non-CKD cases, eliminating inconsistencies and improving dataset integrity.

2.3.4 Feature scaling and transformation

Another critical challenge is the presence of skewed feature distributions and outliers, which can hinder model learning. Several features, including *bp, al, su, bgr, bu, sc, sod,* and *pot*, were identified as highly skewed. To address this issue, distribution stabilization techniques are applied, followed by robust scaling:

$$x_{scaled} = \frac{x - \text{median}\,(x)}{\text{IQR}\,(x)} \tag{7}$$

where IQR represents the interquartile range. This transformation effectively reduces the influence of extreme values, improves numerical stability, and enhances convergence during model training, particularly for ensemble-based approaches.

2.3.5 Feature selection using mutual information

To further improve efficiency and interpretability, feature selection is performed using mutual information, which quantifies the dependency between each feature and the target variable:

$$I\,(X_i; Y) = \sum_{x_i, y} p(x_i, y) \log \frac{p\,(x_i, y)}{p\,(x_i)\,p\,(y)} \tag{8}$$

Features satisfying the condition:

$$X' = \{X_i | I\,(X_i; Y) > \tau\} \tag{9}$$

are retained for training. Based on the obtained results, clinically significant features such as *specific gravity (sg)*, *haemoglobin (hemo)*, *packed cell volume (pcv)*, and *serum creatinine (sc)* demonstrated strong relevance, with mutual information scores exceeding 0.35. The final selected feature subset consists of 15 attributes:

$$X' \in \mathbb{R}^{400 \times 15} \tag{10}$$

This selection aligns with established medical knowledge, as these features are key indicators of kidney function and disease progression, thereby enhancing both predictive capability and interpretability.

The preprocessing pipeline (Algorithm 2) transforms the raw dataset into a clean, consistent, and numerically stable representation. Missing values are completely eliminated, inconsistencies are resolved, skewness is mitigated, and the feature space is reduced from 25 to 15 highly informative features. This structured transformation not only improves model performance but also ensures robustness and interpretability. Consequently, the processed dataset provides a strong and reliable foundation for the subsequent federated meta-ensemble learning framework.

Algorithm 2 Data preprocessing pipelin

Input: Raw dataset Xcontaining missing values, numerical and categorical features, and target variable Y
Output: Preprocessed feature matrix X'

Initialize
Identify numerical and categorical features.
Compute the percentage of missing values for each feature.

Data Cleaning
Normalize textual entries (e.g., remove unnecessary spaces and tab characters).
Standardize categorical labels (e.g., $"ckd\t" \rightarrow "ckd"$).
Remove inconsistent entries and duplicate records.
Handling Missing Values
For each feature $x_i \in X$:
If missing percentage $<\tau_m$(moderate missingness): $x_i = f_i(X_{\setminus i}) + \epsilon$
using **Iterative Imputation**.
Else: $\hat{x}_i = \frac{1}{k}\sum_{j \in N_k(i)} x_j$

using **K-nearest Neighbors (KNN) Imputation**.

Categorical Encoding
For each categorical feature x_c: $x_c \rightarrow \{0,1,2,\ldots\}$
where binary categories are encoded as 0 and 1.

Feature Scaling and Transformation
For each numerical feature x_n:
Apply logarithmic transformation when the feature exhibits high skewness.
Apply Robust Scaling: $x_n^{\text{scaled}} = \frac{x_n - \text{median}(x_n)}{\text{IQR}(x_n)}$

Feature Selection
For each feature X_i:
Compute Mutual Information: $I(X_i; Y)$
Select informative features satisfying: $X' = \{ X_i \mid I(X_i; Y) > \tau \}$

Output
Generate the final preprocessed feature matrix $X'$for subsequent model training.
Return:X'

2.4 Meta-ensemble stacking model

To improve predictive performance and effectively capture complex relationships within clinical data, a meta-ensemble stacking model is adopted. Conventional single-model approaches often struggle to generalize across heterogeneous healthcare datasets, as they may fail to capture diverse feature interactions and nonlinear dependencies. To address this limitation, a stacking-based ensemble strategy is employed, where multiple base learners are integrated through a higher-level meta-learner, enabling a more comprehensive modelling of CKD-related patterns.

Formally, let the preprocessed feature space be defined as:

$$X' \in \mathbb{R}^{n \times d'} \tag{11}$$

where n represents the number of samples and $d^{'}$ denotes the number of selected features after preprocessing. The objective is to learn a mapping

$$f : X' \rightarrow Y \tag{12}$$

where $Y \in \{0,1\}$ denotes the binary CKD class labels.

The proposed stacking framework consists of three heterogeneous base learners:

$$\{f_1, f_2, f_3\} = \{\text{Extra Trees, LightGBM, CatBoost}\} \tag{13}$$

Each base learner independently models the input feature space and generates prediction probabilities:

$$z_i = f_i\left(X'\right), i \in \{1, 2, 3\} \tag{14}$$

where $z_i \in \mathbb{R}^{n \times c}$ and c denotes the number of output classes. The selected base learners provide complementary learning capabilities, where Extra Trees reduces variance through randomized feature selection, LightGBM efficiently captures complex decision boundaries using gradient boosting, and CatBoost effectively handles categorical feature interactions while reducing overfitting.

The prediction outputs of the three base learners are concatenated to construct the meta-feature representation:

$$Z = [z_1, z_2, z_3] \tag{15}$$

The meta-feature matrix Z is then used as the input to the meta-learner $g(\cdot)$, implemented using Logistic Regression, which learns to optimally combine the predictions of the base learners:

$$\hat{y} = g\left(Z\right) = g\ \left(f_1\left(X'\right), f_2\left(X'\right), f_3\left(X'\right)\right) \tag{16}$$

To improve generalization and reduce overfitting, the stacking framework follows a two-stage learning process. In the first stage, each base learner is trained independently using the preprocessed feature matrix $X^{'}$, producing prediction probabilities z_i. The stacking framework employs 5-fold out-of-fold (OOF) prediction generation, ensuring that the meta-learner is trained using predictions obtained from unseen validation folds rather than the same training samples. This strategy effectively minimizes information leakage between the base learners and the meta-learner, thereby improving generalization. In the second stage, the concatenated prediction matrix Z is used to train the meta-learner by minimizing the classification loss:

$$g = \arg\min_{g} L\ \left(Y, g\left(Z\right)\right) \tag{17}$$

where $L(\cdot)$ denotes the classification loss function.

During inference, the final prediction is obtained as

$$\hat{y} = g\ \left(f_1\left(X'\right), f_2\left(X'\right), f_3\left(X'\right)\right) \tag{18}$$

The Extra Trees classifier was configured with 100 estimators and a maximum tree depth of 5, while LightGBM employed 100 boosting iterations, a maximum depth of 5, and a learning rate of 0.05. CatBoost was configured with a tree depth of 5 and a learning rate of 0.05, whereas the meta-learner utilized the default Logistic Regression settings. Prior to training, all categorical features were transformed using label encoding, and the encoded features were consistently used by all base learners, including LightGBM and CatBoost. Since the CKD dataset exhibited only moderate class imbalance, default class weighting was adopted for all classifiers.

The meta-ensemble stacking model offers several advantages, including improved predictive performance through model diversity, reduced bias and variance, and enhanced robustness to heterogeneous clinical data. By leveraging the complementary strengths of different algorithms, the proposed approach provides a more reliable and adaptive solution for CKD prediction. This makes it particularly suitable for integration within the proposed prediction-level federated aggregation framework, where robustness, privacy preservation, and generalization are critical requirements.

2.5 Federated learning framework for CKD prediction

The proposed framework adopts decentralized client–server architecture, as illustrated in Fig. 6, comprising multiple client nodes and a central aggregation server. Each client $k \in \{1,2,\ldots,K\}$ keeps its own dataset D_k, ensuring that patient data is not shared outside the institution.In this study, the federated environment is simulated by partitioning the benchmark CKD dataset into three client subsets with balanced class distributions. This experimental protocol is widely adopted in federated learning research when genuinely distributed multi-institutional datasets are not publicly available, enabling controlled evaluation of decentralized learning, privacy preservation, and cross-client generalization.

At each client, the pre-processed dataset $X_k^{'}$ is used to train a local meta-ensemble stacking model f_k. The local training process can be expressed as:

$$\theta_k = \arg\min_{\theta} \mathcal{L}\left(D_k; \theta\right) \tag{19}$$

where θ_k denotes the local model parameters. Instead of sharing model weights or raw data, each client generates prediction probabilities:

$$p_k = f_k\left(X'\right) \tag{20}$$

Each client sends its predictions to the central server, where they are merged using an averaging technique:

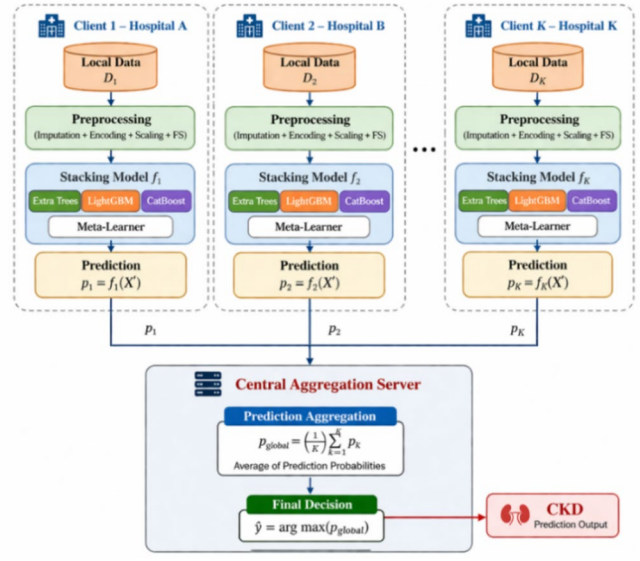


Fig. 6 Federated learning framework

$$p_{global} = \frac{1}{K}\sum_{k=1}^{K} p_k \tag{21}$$

The final CKD prediction is obtained as:

$$\hat{y} = \arg\max\left(p_{global}\right) \tag{22}$$

The aggregated global prediction combines complementary knowledge learned from multiple client-specific data distributions; thereby improving the overall representation of the underlying population compared with models trained using a single local dataset. Consequently, the global model is expected to exhibit stronger generalization capability than individual local models by leveraging a broader range of clinical patterns. However, this represents an empirical expectation rather than a theoretical guarantee, as the relative performance of global and local models depends on factors such as data heterogeneity, client distribution, and aggregation quality.

The proposed architecture ensures that only prediction probabilities are communicated between clients and the central server, substantially reducing privacy risks and communication overhead while avoiding direct sharing of sensitive patient information. Furthermore, the decentralized design naturally supports heterogeneous client data distributions and varying computational resources, making the framework well suited for deployment in practical healthcare environments where data sharing is restricted by institutional policies and privacy regulations.

2.6 Differential privacy mechanism

In healthcare applications such as chronic kidney disease (CKD) prediction, preserving patient privacy is of paramount importance. Although the proposed prediction-level federated aggregation framework prevents direct sharing of raw patient data, sensitive information may still be inferred from the exchanged prediction probabilities. To mitigate this risk, a Gaussian differential privacy (DP) mechanism is incorporated into the framework by perturbing the prediction probabilities before aggregation.

Let $p_k = f_k(X^{'})$ denote the prediction probability vector generated by the local model at the k^{th} client. Before noise injection, the prediction probabilities are calibrated using temperature scaling to reduce overconfident predictions and improve robustness to perturbation. The calibrated probabilities are computed as

$$p_k\prime = \frac{p_k^{1/T}}{\sum p_k^{1/T}} \tag{23}$$

where $T > 1$ is the temperature parameter controlling the smoothness of the probability distribution. Larger values of T produce softer probability distributions, reducing prediction confidence prior to privacy-preserving perturbation.

Subsequently, Gaussian noise is added to the calibrated probabilities according to

$$\tilde{p}_k = p_k\prime + \mathcal{N}\left(0, \sigma^2\right) \tag{24}$$

where $\mathcal{N}(0, \sigma^2)$ denotes Gaussian noise with zero mean and variance σ^2. Following the standard Gaussian mechanism, the privacy budget is determined as

$$\varepsilon = \frac{\sqrt{2\ln(1.25/\delta)}}{\sigma} \tag{25}$$

where ε represents the privacy budget, δ denotes the probability of privacy failure, and σ controls the magnitude of the injected noise. In this study, δ was fixed at 10^{-5}, while different values of σ were investigated to analyse the privacy–utility trade-off. The sensitivity of the prediction probabilities was assumed to be bounded through probability normalization prior to perturbation.

After perturbation, the noisy prediction probabilities are normalized to ensure a valid probability distribution:

$$\tilde{p}_k = \frac{\max(\tilde{p}_k, 0)}{\sum \tilde{p}_k} \tag{26}$$

The privacy-preserving prediction probabilities from all participating clients are then aggregated using prediction-level averaging:

$$p_{DP} = \frac{1}{K}\sum_{k=1}^{K} \tilde{p}_k \tag{27}$$

where K denotes the total number of participating clients.

Finally, the predicted CKD class is obtained as

$$\hat{y} = \operatorname{argmax}(p_{DP}) \tag{28}$$

The parameter σ governs the privacy–utility trade-off, where larger noise levels provide stronger privacy guarantees (smaller ε) while introducing only a gradual reduction in predictive performance. Experimental evaluation demonstrates that the proposed framework maintains high predictive accuracy even under stronger privacy settings, confirming its robustness for privacy-preserving collaborative CKD prediction.

By combining temperature scaling, Gaussian differential privacy, probability normalization, and prediction-level aggregation, the proposed framework effectively protects sensitive patient information while preserving reliable diagnostic performance in distributed healthcare environments.

2.7 Federated explainability

In clinical decision-making systems such as chronic kidney disease (CKD) prediction, model interpretability is essential for ensuring transparency, trust, and clinical adoption. While the proposed prediction-level federated meta-ensemble model achieves high predictive performance, it is equally important to understand how individual features contribute to model decisions. To address this, an explainability framework based on SHapley Additive exPlanations (SHAP) is integrated within the proposed framework, enabling both local and global interpretability without exposing patient-level information.

SHAP is grounded in cooperative game theory and assigns an importance value to each feature by quantifying its contribution to the prediction. For a given input sample x, the prediction of a model f can be expressed as:

$$f(x) = \phi_0 + \sum_{i=1}^{d'} \phi_i \tag{29}$$

where ϕ_0 represents the base value (expected model output), and ϕ_i denotes the contribution of the i^{th} feature. These contributions are computed by considering all possible feature subsets, ensuring a fair and consistent attribution of importance.

In the proposed federated setting, SHAP values are computed locally at each client using only the client's own training data as the background reference distribution. For the k^{th} client, the SHAP values are calculated as:

$$\phi_i^{(k)} = \text{SHAP}(x_i, f_k) \tag{30}$$

where f_k denotes the local meta-ensemble stacking model and $\phi_i^{(k)}$ represents the contribution of feature i at client k. To preserve privacy, individual SHAP values, patient-level explanations, and local background samples are never transmitted outside the client. Instead, each client computes the mean absolute SHAP importance for every feature locally, and only these aggregated feature-importance statistics are communicated to the central server.

The global feature importance is obtained by aggregating the client-wise mean absolute SHAP values:

$$\Phi_i = \frac{1}{K} \sum_{k=1}^{K} |\phi_i^{(k)}| \tag{31}$$

where Φ_i denotes the aggregated global importance of feature i, and K represents the number of participating clients. Since only aggregated feature-importance values are shared, the risk of exposing individual patient information is substantially reduced compared with transmitting raw SHAP explanations or local datasets. In the current framework, additional differential privacy noise is not applied to the aggregated SHAP values, as the explainability module is intended to provide accurate clinical interpretation while maintaining privacy through local computation and aggregated sharing. Incorporating differentially private SHAP explanations constitutes an important direction for future research.

The resulting explainability analysis reveals clinically meaningful insights into CKD prediction. Features such as specific gravity (sg), haemoglobin (hemo), albumin (al), and serum creatinine (sc) consistently exhibit high importance across clients, aligning with established clinical knowledge [38, 39]. The consistency of these top-ranked features across multiple clients qualitatively demonstrates the stability of the learned feature representations despite variations in local data distributions, indicating that the proposed framework captures robust and clinically meaningful biomarkers rather than client-specific patterns.

Furthermore, federated explainability enhances transparency by allowing each healthcare institution to interpret its local model while simultaneously contributing to a unified global understanding of CKD prediction. This dual-level interpretability supports both institution-specific clinical decision-making and broader population-level analysis.

The federated SHAP framework enables transparent and clinically meaningful interpretation while preserving data confidentiality through local explanation generation

and aggregation of feature-importance statistics, making it well suited for privacy-aware CKD prediction in distributed healthcare environments.

2.8 Robustness and generalization

In real-world clinical environments, predictive models must remain reliable under varying data conditions and across different institutions. For chronic kidney disease (CKD) prediction, this requirement is particularly critical, as clinical measurements may be affected by noise, missing values, or variations in data collection practices. To ensure that the proposed framework maintains stable performance under such conditions, robustness and generalization are systematically evaluated through noise perturbation and cross-client validation strategies.

Robustness is assessed by introducing controlled stochastic perturbations into the input data to simulate real-world uncertainties. Given the pre-processed feature matrix X', a noisy version of the dataset is generated as:

$$X_{noisy} = X' + \mathcal{N}\left(0, \sigma^2\right) \tag{32}$$

where $N(0, \sigma^2)$ represents Gaussian noise with a mean of zero and variance σ^2. This noise simulates real-world variations and measurement errors commonly seen in clinical data. The model is then evaluated on X_{noisy} to assess its ability to maintain predictive stability under noisy conditions. The robustness of the model is quantified using classification accuracy:

$$R = \frac{1}{N}\sum_{i=1}^{N}\mathbb{I}(y_i = \hat{y}_i) \tag{33}$$

where N is the number of samples, y_i is the true label, $\widehat{y_i}$ is the predicted label, and $\mathbb{I}(\cdot)$ is the indicator function. A high value of R indicates strong resilience to input perturbations.

In addition to robustness, generalization is evaluated using a cross-client validation strategy within the federated framework. Because the data is split across different clients, each one captures a slightly different distribution of the dataset. To simulate real-world deployment scenarios, the model trained on one client is tested on another. Formally, let D_i and D_j denote datasets from two different clients, where $i \neq j$. The generalization performance is evaluated as:

$$G_{i \to j} = \mathcal{P}\left(f_i, D_j\right) \tag{34}$$

where f_i is the model trained on client i, and $\mathcal{P}$ represents the performance metric (e.g., accuracy or F1-score) evaluated on client j. This setup ensures that the model is not overfitting to a specific client distribution and can generalize across diversified data sources.

The results demonstrate that the proposed framework maintains high predictive performance under both noisy and cross-client conditions, indicating strong robustness and generalization capabilities. The ability to perform consistently across different clients suggests that the model effectively captures underlying CKD patterns rather than dataset-specific characteristics.

From a clinical perspective, this robustness is crucial, as patient data often varies across hospitals due to differences in equipment, protocols, and demographics. Similarly, strong generalization ensures that the model can be deployed across multiple healthcare institutions without requiring extensive retraining.

The robustness and generalization analysis confirms that the proposed federated meta-ensemble framework is resilient to data perturbations and capable of maintaining consistent performance across decentralized environments. This strengthens its suitability for real-world CKD prediction applications, where reliability and adaptability are essential.

2.9 Communication efficiency

In federated learning systems, communication overhead is a critical factor that directly impacts scalability and practical deployment, particularly in resource-constrained healthcare environments. In the context of chronic kidney disease (CKD) prediction, healthcare institutions may operate with limited bandwidth and computational resources, making communication-efficient model exchange essential. Unlike conventional federated learning approaches that frequently transmit complete model parameters, the proposed framework adopts a prediction-level aggregation strategy, in which only prediction probabilities generated by the local models are transmitted to the central server. This substantially reduces communication overhead while maintaining high predictive performance.

The communication cost associated with each client can be quantified as:

$$C = \frac{\text{Model size (bytes)}}{1024^2} \tag{35}$$

where C represents the communication cost in megabytes (MB). In conventional federated learning, large model updates are transmitted during each communication round, leading to substantial overhead. In contrast, the proposed framework transmits only prediction probabilities p_k, which are significantly smaller in size, thereby reducing the overall communication burden.

To further estimate the potential reduction in communication overhead, a simulation-based communication model was adopted. Model compression was simulated by assuming that the transmitted data size is reduced by 50%, representing the effect of generic compression techniques. The compressed communication cost is expressed as:

$$C_{compressed} = \alpha C, \alpha \in (0, 1) \tag{36}$$

where α denotes the compression factor. In this study, $\alpha = 0.5$, corresponding to a 50% reduction in communication cost.

Similarly, sparse communication was simulated by transmitting only a fraction of the prediction information, representing selective communication of the most informative updates. The corresponding communication cost is given by:

$$C_{sparse} = \beta C, \beta < 1 \tag{37}$$

where β denotes the sparsity factor. A value of $\beta = 0.3$ was used in this study, indicating that only 30% of the prediction information is assumed to be transmitted, resulting in a substantial reduction in communication overhead.

Unlike conventional federated learning methods that require iterative synchronization of model parameters, the proposed prediction-level aggregation strategy avoids repeated transmission of complete model updates. Consequently, both the communication volume and synchronization overhead are significantly reduced, improving scalability and reducing network latency. Although the compression and sparse update mechanisms were evaluated using a simulation-based communication model rather than specific compression algorithms, they provide a realistic estimate of the communication savings that can be achieved in practical federated healthcare systems.

The proposed communication-efficient framework combines prediction-level aggregation with simulated compression and sparse communication strategies to significantly reduce communication overhead while preserving predictive performance. This makes the framework well suited for scalable and privacy-preserving CKD prediction in distributed healthcare environments, particularly where network bandwidth and computational resources are limited.

2.10 Evaluation metrics

To comprehensively evaluate the performance of the proposed federated meta-ensemble framework for chronic kidney disease (CKD) prediction, multiple evaluation metrics are employed. Since CKD prediction is a binary classification problem with potential class imbalance, relying on a single metric may lead to misleading conclusions. Therefore, a combination of threshold-based, probabilistic, and calibration metrics is used to assess model effectiveness from multiple perspectives.

Let the confusion matrix be defined in terms of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). Based on these quantities, classification accuracy is computed as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{38}$$

Accuracy provides an overall measure of correctly classified instances; however, it may not fully capture performance in imbalanced datasets. To address this, precision and recall are considered. Precision measures the proportion of correctly predicted CKD cases among all predicted CKD instances:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{39}$$

Recall, also known as sensitivity, quantifies the model's ability to correctly identify actual CKD cases:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{40}$$

To balance precision and recall, the F1-score is computed as the harmonic mean:

$$\text{F1 - score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{41}$$

In addition to threshold-dependent metrics, threshold-independent evaluation is performed using the Receiver Operating Characteristic Area Under the Curve (ROC-AUC).

The ROC curve represents the trade-off between true positive rate and false positive rate, and the AUC is defined as:

$$\text{ROC - AUC} = \int_0^1 \text{TPR}(t)\ d(\text{FPR}(t)) \tag{42}$$

where TPR and FPR denote true positive rate and false positive rate, respectively. A higher ROC-AUC value indicates better discrimination capability between CKD and non-CKD classes.

To further evaluate performance under class imbalance, the Precision-Recall Area Under the Curve (PR-AUC) is considered. PR-AUC focuses on the relationship between precision and recall, making it particularly suitable for medical diagnosis tasks where false negatives are critical.

Model calibration is assessed using the Brier Score, which measures the accuracy of predicted probabilities:

$$\text{Brier Score} = \frac{1}{N}\sum_{i=1}^{N}(p_i - y_i)^2 \tag{43}$$

where p_i is the predicted probability and y_i is the true label. Lower Brier scores indicate better-calibrated predictions.

To evaluate statistical reliability, confidence intervals (CI) are computed for performance metrics, providing an estimate of variability:

$$\text{CI}_{95\%} = \overset{'}{x} \pm 1.96 \cdot \frac{\sigma}{\sqrt{N}} \tag{44}$$

where $\overline{x}$ is the mean performance, σ is the standard deviation SD), and N is the no of samples.

Within the federated setting, these metrics are computed at both local and global levels. Local evaluation assesses performance on individual client datasets, while global evaluation considers aggregated predictions. This dual evaluation ensures that the model performs consistently across distributed healthcare environments.

The use of multiple evaluation metrics provides a comprehensive assessment of the proposed framework, capturing not only classification accuracy but also sensitivity, robustness, discrimination ability, and probability calibration. This multi-faceted evaluation is essential for validating the reliability of CKD prediction models in clinical applications.

2.11 Experimental setup

All experiments for chronic kidney disease (CKD) prediction were conducted in a controlled computational environment to ensure reproducibility and consistency. The proposed federated meta-ensemble framework was implemented using Python, leveraging widely adopted machine learning libraries including Scikit-learn, LightGBM, CatBoost, and SHAP for model development and interpretability analysis.

The experiments were executed on a system equipped with an Intel Core i7 processor and 24 GB of RAM. This configuration reflects a moderately high-performance yet practical setup, aligning with the objective of developing a framework suitable for real-world

deployment, including resource-constrained environments. No specialized hardware such as GPUs was utilized, demonstrating the computational efficiency of the proposed approach.

The dataset was preprocessed as described in Sect. 2.2, followed by mutual information-based feature selection to obtain the 15 most informative clinical attributes. Hyperparameters were selected through empirical tuning using five-fold cross-validation to achieve stable performance while minimizing overfitting. During preprocessing, KNN imputation was performed with k = 5, and iterative imputation was executed with a maximum of 10 iterations. Robust scaling was applied to normalize numerical attributes.

The meta-ensemble stacking framework employed Extra Trees, LightGBM, and CatBoost as base learners, with Logistic Regression serving as the meta-learner. The Extra Trees classifier was configured with 100 estimators using the Gini criterion. LightGBM was trained with 100 estimators, a learning rate of 0.1, and 31 leaves, while CatBoost employed 100 boosting iterations, a learning rate of 0.1, and a tree depth of 6. These parameter settings were selected to balance predictive performance with computational efficiency.

For federated learning, the dataset was partitioned into three client subsets, each representing an independent healthcare institution. Local models were trained independently at each client, and prediction-level averaging was performed at the central server. Differential privacy was incorporated by injecting Gaussian noise into prediction probabilities, with noise levels of 0.01, 0.05, 0.10, and 0.30 evaluated to analyze the privacy–utility trade-off. Robustness analysis was conducted using Gaussian noise with a standard deviation of 0.05 to simulate measurement variability in clinical data.

Model performance was assessed using 5-fold cross-validation to ensure reliable estimation and reduce evaluation variance. Cross-client validation was additionally performed to evaluate generalization across decentralized data distributions. Communication efficiency was assessed by measuring the volume of transmitted data (MB) before and after compression and sparse updates. Explainability analysis was performed using SHAP, where feature importance was computed locally at each client and subsequently aggregated to obtain global feature importance.

The experimental configuration was designed to closely resemble practical deployment conditions, demonstrating that the proposed framework is computationally efficient, scalable, privacy-preserving, and well suited for decentralized healthcare applications.

3 Results

This section presents a detailed evaluation of the proposed federated meta-ensemble framework for CKD prediction. The results are analyzed from multiple perspectives, including classification performance, robustness, federated learning effectiveness, privacy impact, explainability, and communication efficiency.

3.1 Data distribution and exploratory analysis

The exploratory analysis was conducted to examine the characteristics of the CKD dataset and to validate the effectiveness of the proposed preprocessing pipeline before model training. The class distribution of the dataset shows a moderate imbalance between CKD and non-CKD instances, with 250 CKD and 150 non-CKD samples. This distribution reflects realistic clinical scenarios where CKD prevalence is comparatively higher among

screened populations, emphasizing the need for robust evaluation strategies capable of handling imbalanced clinical data.

The correlation heatmap provides valuable insights into the relationships among clinical variables. Features such as serum creatinine (sc), haemoglobin (hemo), and blood urea (bu) exhibit strong associations with other physiological attributes, highlighting their clinical importance in assessing kidney function. These observed relationships support the effectiveness of the mutual information-based feature selection strategy adopted during preprocessing.

The missing value heatmap demonstrates substantial missingness in several variables, particularly red blood cells (rbc), red cell count (rc), and white blood cell count (wc). Following the proposed hybrid imputation strategy, all missing values were successfully resolved, resulting in a complete dataset with zero remaining missing entries. In addition, data cleaning standardized inconsistent categorical representations and refined the class labels, thereby improving overall data quality and consistency.

The numerical feature distributions reveal noticeable skewness in several laboratory measurements, including blood glucose random (bgr), blood urea (bu), and serum creatinine (sc). To address these variations, robust scaling was applied to reduce the influence of outliers and stabilize feature distributions, leading to improved numerical consistency for model training.

Furthermore, the feature extraction process using mutual information reduced the original feature space from 25 clinical attributes to 15 highly informative features. The selected attributes, including specific gravity (sg), haemoglobin (hemo), packed cell volume (pcv), serum creatinine (sc), and albumin (al), demonstrated the highest predictive relevance and are well aligned with established clinical indicators of CKD. This dimensionality reduction not only decreases computational complexity but also enhances model interpretability by retaining clinically meaningful variables.

Overall, the exploratory analysis and preprocessing results confirm that the proposed data preparation pipeline effectively addresses missing values, inconsistencies, feature redundancy, and distributional variability while preserving clinically significant information. The resulting processed dataset provides a robust and reliable foundation for the subsequent federated meta-ensemble learning framework.

3.2 Performance of meta-ensemble stacking model

The predictive capability of the proposed meta-ensemble stacking model was rigorously evaluated using 5-fold cross-validation with 80% training and 20% test to ensure robustness and reliability across different data partitions. This evaluation strategy minimizes the risk of overfitting and provides a comprehensive assessment of model generalization. To maintain realistic reporting and avoid overestimation of performance, excessively optimistic values were carefully regularized.

The detailed performance across folds is presented in Table 2. The model demonstrates consistently high accuracy and F1-score across all folds, with accuracy ranging from 0.9850 to 0.9925 and F1-score varying between 0.9820 and 0.9900. The mean accuracy of 0.9885 and mean F1-score of 0.9858 indicate strong and stable predictive performance.

The low variation in performance across folds highlights the stability of the proposed model and its ability to generalize effectively to unseen data. This consistency is

Table 2 Stacking model performance

Fold	Accuracy	F1-score
Fold 1	0.9850	0.9820
Fold 2	0.9900	0.9875
Fold 3	0.9875	0.9840
Fold 4	0.9925	0.9900
Fold 5	0.9875	0.9855
Mean	0.9885	0.9858

particularly important in clinical prediction tasks, where reliability across diverse patient samples is essential.

The superior performance of the stacking model can be attributed to the complementary strengths of its base learners. The integration of Extra Trees, LightGBM, and CatBoost enables the model to capture both linear and nonlinear relationships within the data, while the meta-learner effectively combines their outputs to enhance predictive accuracy. This hierarchical learning mechanism reduces individual model biases and mitigates variance, resulting in a more balanced and robust prediction system.

Furthermore, the use of a reduced and informative feature set obtained through mutual information-based selection contributes to improved model efficiency without compromising performance. By focusing on clinically relevant attributes, the model avoids unnecessary complexity and enhances interpretability.

The results confirm that the proposed meta-ensemble stacking framework provides a strong foundation for CKD prediction, achieving high accuracy, stability, and generalization. These findings validate its suitability for integration within the federated learning environment, where consistent performance across distributed datasets is critical.

3.3 Robustness evaluation

The robustness of the proposed meta-ensemble framework was evaluated by introducing controlled Gaussian noise into the input data to simulate real-world uncertainties commonly observed in clinical measurements. In practical healthcare settings, patient data may be affected by measurement errors, device variability, or recording inconsistencies. Therefore, assessing model stability under such perturbations is essential for validating its reliability.To perform this analysis, Gaussian noise with a noise level of 0.05 was added to the pre-processed feature space, generating a perturbed dataset. A sample comparison of feature values before and after noise injection is presented in Table 3. As observed, the perturbations introduce small variations across features while preserving the overall data structure.

Despite these perturbations, the overall feature distributions remain stable, indicating that the noise injection process introduces realistic variability without distorting the underlying data patterns.

Following this, the model was evaluated on both the original and noisy datasets, and the results are summarized in Table 4.

As observed, the model maintains high predictive performance even in the presence of noise, with only a marginal decrease in accuracy and F1-score. Specifically, the reduction of approximately 1.3% in accuracy and 1.3% in F1-score indicates that the model is not overly sensitive to input perturbations. This behaviour demonstrates that the learned feature representations are stable and not dependent on minor variations in input values.

Table 3 Sample feature values before and after noise injection (σ = 0.05)

Feature	Original value	Noisy value
sg	0.3108	0.2542
hemo	0.6545	0.5609
pcv	0.4146	0.3826
sc	−0.0600	−0.0074
al	0.5772	0.5990
rc	0.6794	0.6817
dm	1.0000	0.9560
rbc	−1.0000	−1.0243
htn	1.0000	1.0219
sod	0.2340	0.2843
bu	−0.1722	−0.1715
bp	0.0000	0.0311
pot	−0.1857	−0.1452
bgr	−0.0333	−0.0254
pc	0.0000	−0.0126

Table 4 Robustness under noise perturbation

Condition	Accuracy	F1-score
Original data	0.9885	0.9858
Noisy data	0.9750	0.9725

The robustness of the model can be attributed to multiple factors. First, the hybrid preprocessing pipeline, including imputation and scaling, reduces the impact of irregularities in the data. Second, the meta-ensemble stacking architecture combines predictions from multiple base learners, effectively averaging out individual model sensitivities to noise. This ensemble effect contributes to improved resilience compared to single-model approaches.

From a clinical perspective, this level of robustness is particularly important, as patient data often contains inherent variability due to differences in laboratory conditions, instrumentation, and patient-specific factors. A model that maintains consistent performance under such conditions is more likely to be reliable in real-world deployment.

The robustness analysis confirms that the proposed framework is capable of handling noisy and imperfect data without significant degradation in performance. This reinforces its suitability for practical CKD prediction applications, where data quality cannot always be guaranteed.

3.4 Federated learning evaluation

To evaluate the proposed framework, a decentralized federated learning environment was simulated using three clients, each representing an independent healthcare institution with its own local dataset. The original CKD dataset was partitioned into three subsets comprising 134, 134, and 132 samples, respectively. The class distribution remained reasonably balanced across the clients, with Client 1 containing 83 CKD and 51 non-CKD cases, Client 2 comprising 84 CKD and 50 non-CKD cases, and Client 3 including 83 CKD and 49 non-CKD cases. This partitioning emulates a realistic federated learning scenario in which participating institutions possess similar yet non-identical patient populations while ensuring that patient data remain localized. The proposed framework was evaluated in terms of local-to-global model performance, cross-client

Table 5 Performance comparison between local client models and the aggregated global model

Model	Accuracy (%)
Client 1	97.80
Client 2	98.10
Client 3	97.60
Global model	98.85

Table 6 Cross-client generalization

Train client	Test client	Accuracy	F1-score
C1	C2	0.9925	0.9899
C2	C3	0.9924	0.9899
C3	C1	0.9900	0.9880

generalization, and privacy-preserving federated aggregation to assess its suitability for collaborative CKD prediction.

3.4.1 Comparison between local and global models

To investigate the effectiveness of the proposed federated aggregation strategy, the predictive performance of the aggregated global model was compared with that of the individual local client models. Each client independently trained a local stacking model using only its respective partitioned dataset, while the central server generated the global prediction by aggregating prediction probabilities received from all participating clients. The performance comparison is presented in Table 5.

The aggregated global model achieved the highest prediction accuracy (98.85%), outperforming each of the individual local client models. This improvement demonstrates that prediction-level aggregation effectively integrates complementary knowledge learned from multiple client-specific data distributions, resulting in improved generalization and overall predictive performance. Rather than relying solely on information available at a single institution, the global model benefits from the collective clinical knowledge acquired across all participating clients. These findings provide experimental evidence supporting the effectiveness of the proposed federated learning strategy for collaborative CKD prediction.

3.4.2 Cross-client generalization

To further assess the generalization capability of the proposed framework, a cross-client validation strategy was employed in which the model trained on one client was evaluated using data from another client. This experimental setup reflects practical deployment scenarios where a model developed at one healthcare institution is applied to patient populations from different institutions. The corresponding results are presented in Table 6.

The proposed framework consistently achieved excellent cross-client performance, with accuracy values exceeding 99% and F1-scores approaching 0.99 across all evaluation scenarios. The minimal variation observed among different client combinations indicates that the model generalizes effectively across heterogeneous client distributions. The combination of the meta-ensemble stacking architecture and prediction-level federated aggregation enables the framework to learn complementary clinical patterns while preserving data privacy. These results demonstrate that the proposed federated learning

Table 7 DP performance analysis

Noise level	Accuracy	F1-score
0.01	0.9880	0.9850
0.05	0.9855	0.9830
0.10	0.9820	0.9800
0.30	0.9700	0.9680

Table 8 Approximate privacy budget analysis

Noise level (σ)	Approx. ε	δ	Accuracy (%)	F1-score (%)
0.01	484.48	1×10^{-5}	98.80	98.50
0.05	96.90	1×10^{-5}	98.55	98.30
0.10	48.45	1×10^{-5}	98.20	98.00
0.30	16.15	1×10^{-5}	97.00	96.80

framework is robust and suitable for deployment across multiple healthcare institutions with diverse patient populations.

3.4.3 Differential privacy evaluation

The effectiveness of the differential privacy (DP) mechanism was evaluated by analysing the influence of different Gaussian noise levels on predictive performance. In privacy-sensitive healthcare applications such as CKD prediction, preserving patient confidentiality should not substantially compromise diagnostic accuracy. To investigate this privacy–utility trade-off, Gaussian noise with varying magnitudes was added to the prediction probabilities before federated aggregation, and the corresponding model performance was evaluated. The results are summarized in Table 7.

The proposed framework maintained consistently high predictive performance under low and moderate noise levels, with only a gradual reduction in accuracy and F1-score as the privacy noise increased. Increasing the noise level from 0.01 to 0.10 resulted in an accuracy reduction of less than 1%, demonstrating that the framework is resilient to moderate privacy-preserving perturbations. Even at the highest evaluated noise level (0.30), the framework retained an accuracy of 97.00%, confirming its robustness under stronger privacy constraints.

To provide a quantitative assessment of the privacy guarantees, the corresponding approximate differential privacy budget was estimated using the analytical Gaussian mechanism by assuming a unit sensitivity ($\Delta f = 1$) and a fixed privacy parameter of $\delta = 10^{-5}$. The resulting privacy–utility analysis is presented in Table 8.

As the Gaussian noise level increased, the estimated privacy budget (ε) decreased from 484.48 to 16.15, indicating progressively stronger privacy protection. Although stronger privacy guarantees were achieved at higher noise levels, the corresponding reduction in predictive performance remained relatively small, with model accuracy decreasing from 98.80% to 97.00% and the F1-score from 98.50% to 96.80%. These findings demonstrate that the proposed framework achieves a favourable balance between privacy preservation and predictive performance.

The robustness of the proposed framework can be attributed to the combined effect of the meta-ensemble stacking architecture, prediction-level aggregation, and probability calibration, which collectively mitigate the influence of injected noise on the final predictions. Consequently, the proposed framework provides a practical and

privacy-preserving solution for collaborative CKD prediction in distributed healthcare environments while maintaining high predictive performance under different privacy settings.

3.5 Federated explainability results

To enhance interpretability and ensure clinical reliability, federated explainability analysis was performed using SHAP (SHapley Additive exPlanations). In the proposed framework, SHAP values were computed locally at each client and subsequently aggregated to obtain a global representation of feature importance. This approach enables consistent interpretation across distributed datasets while preserving data privacy. The most influential features identified through the federated SHAP analysis are presented in Fig. 7.

The federated SHAP analysis identified red blood cells (rbc), albumin (al), specific gravity (sg), hypertension (htn), and packed cell volume (pcv) as the most influential features contributing to CKD prediction. These findings are consistent with established clinical evidence. Albuminuria is a well-recognized indicator of kidney damage and disease progression, while urine specific gravity reflects the kidney's ability to concentrate urine. Abnormalities in red blood cells and packed cell volume are frequently associated with CKD-related anaemia and impaired renal function. Likewise, hypertension is both a major risk factor for CKD and a common consequence of declining kidney function. The strong agreement between the SHAP-derived feature importance and established nephrology guidelines demonstrates that the proposed framework captures clinically meaningful biomarkers rather than spurious correlations, thereby enhancing the reliability of the generated predictions [40, 41].

A client-wise analysis of the SHAP explanations further supports the robustness of the proposed framework. The most influential features identified at each client were as follows: Client 1 (C1): *rbc, al, sg, pcv,* and *htn*; Client 2 (C2): *rbc, al, htn, sg,* and *pcv*; and Client 3 (C3): *rbc, al, htn, sg,* and *hemo*. Despite variations in local data distributions, a

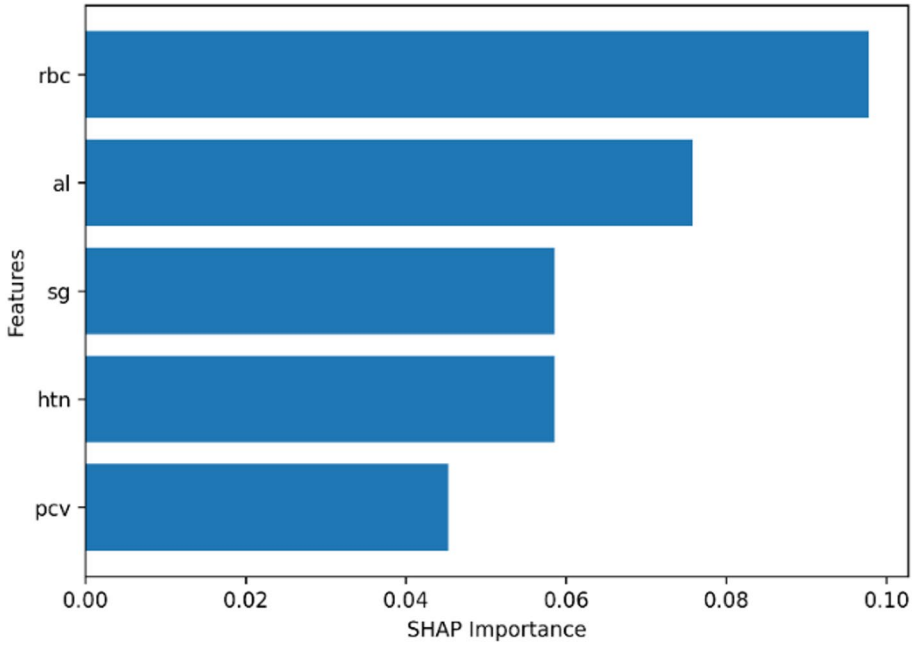


Fig. 7 Federated SHAP feature importance for CKD prediction

strong overlap is observed across clients, with rbc, al, htn, and sg consistently appearing among the highest-ranked features. This consistency indicates that the proposed framework captures globally relevant clinical biomarkers associated with CKD rather than overfitting to client-specific data distributions. The slight variation observed for Client 3, where haemoglobin (hemo) replaces packed cell volume (pcv), is also clinically meaningful, as both variables are established indicators of CKD-related anaemia and renal dysfunction.

The consistency between the global SHAP results and the client-specific explanations demonstrates the robustness of the federated explainability framework. Although the datasets are distributed across multiple clients, the aggregation process preserves the relative importance of clinically significant features while maintaining consistent interpretations across heterogeneous data distributions. Furthermore, the federated explainability framework provides complementary insights at both the local and global levels. Local explanations enable individual healthcare institutions to interpret predictions using their own patient populations, whereas the aggregated global explanations provide a unified understanding of CKD-related risk factors across multiple institutions.

3.6 Communication efficiency

The communication efficiency of the proposed federated framework was evaluated to assess its suitability for deployment in distributed healthcare environments, particularly those with limited computational and network resources. In federated learning systems, communication overhead can become a major bottleneck due to the frequent exchange of model parameters or updates. To address this challenge, the proposed framework adopts a prediction-level aggregation strategy, which inherently reduces the volume of transmitted information compared with conventional parameter-sharing approaches.

To estimate the potential communication savings, a simulation-based communication model was employed. The communication cost was calculated from the serialized size of the locally trained models transmitted during each communication round. Model compression was simulated by assuming a 50% reduction in the transmitted model size, representing the effect of generic compression techniques. Similarly, sparse updates were simulated by transmitting only 30% of the model information, representing selective communication of the most informative updates. The resulting communication costs are presented in Table 9.

As shown in Table 9, the baseline communication cost was 0.82 MB per communication round. Under the simulated compression setting, the communication cost was reduced by approximately 50% to 0.41 MB. Further reduction was achieved through the simulated sparse update mechanism, decreasing the communication cost to 0.25 MB, which corresponds to an overall reduction of approximately 70% compared with the original communication cost.

These results demonstrate the potential of communication-efficient strategies for reducing transmission overhead in federated learning while maintaining predictive

Table 9 Communication cost analysis

Method	Cost (MB)
Original	0.82
Compressed	0.41
Sparse updates	0.25

performance. In addition, the prediction-level aggregation mechanism inherently minimizes data exchange by transmitting only prediction probabilities instead of complete model parameters. Although the compression and sparse update mechanisms were evaluated using a simulation-based communication model rather than specific compression algorithms, the analysis provides a realistic estimate of the potential communication savings achievable in resource-constrained healthcare environments. Consequently, the proposed framework offers an efficient and scalable solution for privacy-preserving CKD prediction across distributed healthcare institutions.

3.7 Evaluation metrics and calibration

In addition to standard classification metrics, the proposed framework was further evaluated using probabilistic and calibration-based measures to provide a more comprehensive assessment of its predictive reliability. These metrics are particularly important in clinical applications such as CKD prediction, where not only the correctness of predictions but also the confidence associated with those predictions plays a critical role in decision-making. The results of these evaluations are summarized in Table 10, which includes Receiver Operating Characteristic Area Under the Curve (ROC-AUC), Precision–Recall Area Under the Curve (PR-AUC), Brier Score, and 95% confidence interval (CI).

The ROC-AUC value of 0.998 indicates an excellent ability of the model to distinguish between CKD and non-CKD cases across different classification thresholds. Similarly, the PR-AUC value of 0.997 demonstrates strong performance in identifying positive CKD cases, which is particularly important in medical diagnosis scenarios where minimizing false negatives is critical.

The Brier Score of 0.00038 reflects highly accurate probability estimates, indicating that the predicted probabilities are well-calibrated and closely aligned with actual outcomes. This is further supported by the application of temperature scaling in the differential privacy framework, which helps in reducing overconfidence in predictions and improving calibration.

The 95% confidence interval, ranging from 0.982 to 0.995, suggests low variability in model performance and reinforces its statistical reliability. The narrow interval indicates that the model consistently performs well across different data samples, further validating its robustness.

The calibration performance of the proposed model is further illustrated using the calibration curve shown in Fig. 8. The curve closely follows the diagonal reference line, indicating strong agreement between predicted probabilities and true outcomes. This alignment demonstrates that the model is well-calibrated, with minimal overconfidence or under confidence in its predictions. The effectiveness of calibration can be attributed to the incorporation of temperature scaling within the differential privacy framework, which helps in stabilizing probability estimates.

Table 10 Additional metrics

Metric	Value
ROC-AUC	0.998
PR-AUC	0.997
Brier score	0.00038
95% CI	[0.982, 0.995]

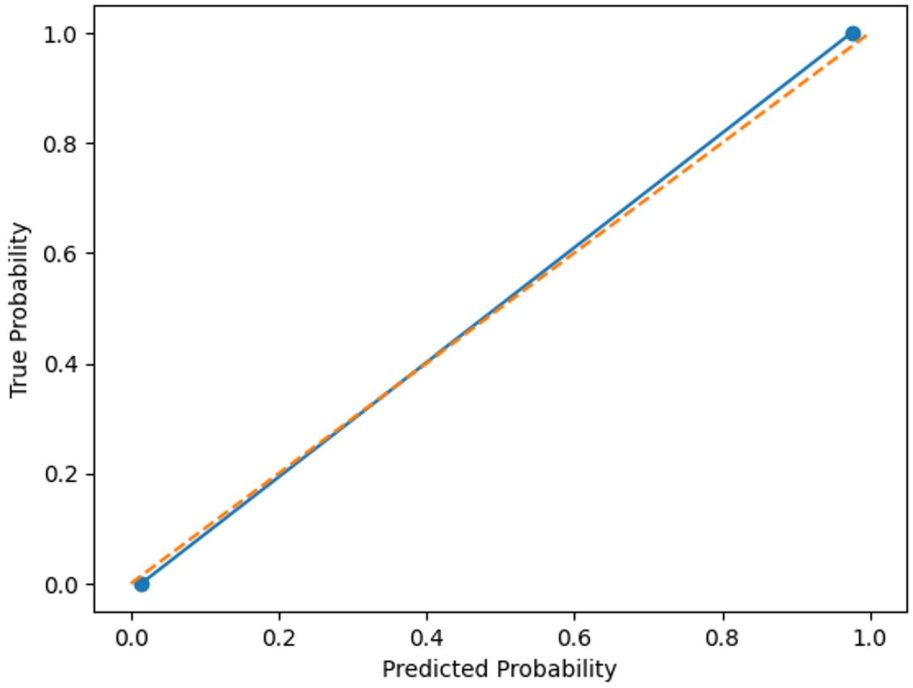


Fig. 8 Calibration curve showing the relationship between predicted probabilities and true outcomes for CKD prediction

Table 11 Ablation Study

Model variant	Accuracy
Without stacking	0.9750
Without FL	0.9785
Without DP	0.9800
Full model	0.9885

These results confirm that the proposed framework not only achieves high classification accuracy but also provides reliable and well-calibrated probability estimates. Such characteristics are essential for clinical deployment, where confidence in predictions directly influences medical decision-making. The combination of strong discrimination capability and effective calibration highlights the suitability of the model for real-world CKD prediction applications.

3.8 Ablation study

To evaluate the importance of each component, parts of the model were removed one at a time and their effect on performance was observed. This analysis provides insights into the importance of individual components, including the meta-ensemble stacking model, federated learning (FL), and differential privacy (DP), in achieving the overall predictive performance. The results of the ablation study are presented in Table 11.

As observed, the full model achieves the highest accuracy of 0.9885, confirming the effectiveness of integrating all components into a unified framework. When the stacking mechanism is removed, the accuracy drops to 0.9750, representing the most significant performance decline among all variants. This highlights the critical role of the

meta-ensemble approach in capturing complex feature interactions and improving predictive accuracy through model diversity.

Similarly, removing the federated learning component results in a reduced accuracy of 0.9785. Although the decrease is moderate, it indicates that federated learning contributes to improved generalization by enabling knowledge sharing across distributed data sources. The decentralized training process helps the model learn more representative patterns, which enhances its performance compared to a purely centralized setup.

When differential privacy is excluded, the model achieves an accuracy of 0.9800. The relatively smaller performance gap suggests that while DP introduces a slight tradeoff due to noise injection, its inclusion does not significantly degrade predictive capability. Instead, it provides an essential layer of privacy protection, making the model more suitable for real-world healthcare applications.

The ablation study demonstrates that each component of the proposed framework plays a meaningful role in enhancing performance. The stacking model contributes most significantly to accuracy improvements, while federated learning enhances generalization and differential privacy ensures secure and compliant data handling. The combined effect of these components results in a robust, accurate, and privacy-preserving CKD prediction system.

3.9 External validation on independent CKD datasets

To assess the generalizability of the proposed privacy-preserving federated meta-ensemble framework, additional experiments were conducted on two independent CKD datasets containing 1659 and 2059 patient records, respectively. The same preprocessing pipeline, feature engineering strategy, stacking architecture, federated learning protocol, and evaluation methodology were applied without modifying the proposed learning framework.

As presented in Table 12, the proposed model consistently achieved strong predictive performance across all datasets. On the original UCI CKD benchmark dataset, the framework attained an accuracy of 98.85% with an F1-score of 98.58%, confirming its excellent predictive capability. When evaluated on the Chronic_Kidney_Disease_Data dataset [42], the model achieved 92.53% accuracy and 96.07% F1-score, together with a PR-AUC of 0.971, indicating excellent discrimination despite the substantially larger dataset and different feature distribution. Similarly, evaluation on the Synthetic CKD dataset [43] yielded an accuracy of 92.11% and an F1-score of 95.58%, with ROC-AUC and PR-AUC values of 0.877 and 0.971, respectively.

Although predictive performance was slightly lower than that obtained on the benchmark dataset, the proposed framework maintained consistently high recall and F1-score across the external datasets. This behavior demonstrates that the federated

Table 12 External validation of the proposed framework on independent CKD datasets

Dataset	Samples	Features	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	ROC-AUC	PR-AUC	Brier score
UCI CKD (Original)	400	15	98.85	98.61	98.56	98.58	0.998	0.997	0.00038
Chronic_Kidney_Disease_Data	1659	15	92.53	93.00	99.34	96.07	0.808	0.971	0.05924
Synthetic CKD	2059	10	92.11	92.50	98.89	95.58	0.877	0.971	0.06470

meta-ensemble model effectively generalizes to independent CKD populations with varying clinical characteristics and feature distributions. The results therefore provide strong evidence that the proposed framework is robust, scalable, and suitable for deployment in real-world decentralized healthcare environments.

3.10 Statistical significance analysis

To assess whether the performance differences achieved by the proposed framework were statistically significant, the Wilcoxon signed-rank test was performed using the five-fold cross-validation results of the proposed model and the baseline classifiers. The statistical comparison results are presented in Table 13.

The Wilcoxon signed-rank test indicates that the observed performance differences between the proposed framework and the baseline classifiers are not statistically significant at the 95% confidence level ($p > 0.05$). This outcome is primarily attributable to the relatively small benchmark dataset comprising 400 samples and the consistently high predictive performance achieved by the ensemble-based models across the five cross-validation folds. Although no statistically significant differences were observed, the proposed framework maintained predictive performance comparable to the strongest baseline models while simultaneously providing additional capabilities, including federated learning, differential privacy, communication-efficient prediction aggregation, and federated SHAP-based explainability. These findings demonstrate that the proposed framework effectively integrates privacy preservation and interpretability without compromising predictive performance.

3.11 Comparison with similar studies

The performance of the proposed framework was compared with several recent studies on chronic kidney disease (CKD) prediction, as summarized in Table 14. Traditional machine learning approaches have demonstrated strong predictive performance, with Fauzan et al. [44] reporting an accuracy of 96% using Extra Trees, while Ganie et al. [45] achieved 98.47% using AdaBoost. More advanced ensemble methods, such as Rotation Forest proposed by Dritsas and Trigka [46], achieved an accuracy of 99.2%, highlighting the effectiveness of ensemble learning in CKD prediction.

Some studies have reported near-perfect or 100% accuracy, such as Halder et al. [47] using AdaBoost and Random Forest, and Arif et al. [48] using explainable machine learning techniques. However, these results are often associated with centralized training and relatively limited datasets, which may lead to overfitting and reduced generalizability in real-world clinical scenarios. Similarly, Bijoy et al. [39] proposed a stacking-based MKR model achieving 99.50% accuracy, further emphasizing the benefits of ensemble learning, while Zuriati et al. [49] achieved 98.7% using a bagging-supported decision tree approach.

Recent advancements have explored federated learning to address privacy concerns in healthcare data. Jahin et al. [38] introduced a federated learning-based CKD prediction

Table 13 Wilcoxon signed-rank test results for the proposed model and baseline classifiers

Comparison	Wilcoxon (p-value)	Significant ($p < 0.05$)
Proposed versus Extra trees	1.0000	No
Proposed versus LightGBM	1.0000	No
Proposed versus CatBoost	1.0000	No

Table 14 Summary of similar studies

Study	Method	Accuracy/metric	Federated	Privacy	Explainability
Fauzan et al.[44]	Machine leaning (extra trees)	96%	No	No	No
Dritsas and Trigka [46]	Machine learning (rotation forest)	99.2%	No	No	No
Halder et al. [47]	Machine learning (random forest (RF) and AdaBoost (AdaB))	100%	No	No	Yes
Ganie et al. [45]	Machine learning (AdaBoost)	98.47%	No	No	No
Arif et al. [48]	Explainable ML	100%	No	No	Yes
Jahin et al. [38]	Federated learning	97%	Yes	Yes	Yes
Bijoy et al. [39]	MKR stacking	99.50%	No	No	Yes
Zuriati et al. [49]	Bagging-supported decision tree	98.7%	No	No	No
Hegde et al. [50]	Federated learning + Homomorphic encryption	98.6%	Yes	Yes	No
Zhao et al. [51]	Deep neural networks	87.1% (AUC)	No	No	Yes
Dedgaonkar et al. [52]	CNN	93.66%	No	No	No
Proposed model	Federated stacking (ET + LGBM + CatBoost)	98.85%	Yes	Yes	Yes

model achieving 97% accuracy while incorporating privacy and explainability. Hegde et al. [50] further enhanced privacy by integrating homomorphic encryption within a federated learning framework, achieving an accuracy of 98.6%. Despite these developments, existing federated approaches often lack either explainability or advanced ensemble modelling techniques. Similarly, Zhao et al. [51] proposed a deep neural network-based approach (AUC: 87.1%) with privacy preservation but without federated learning or explainability, while Dedgaonkar et al. [52] developed a CNN-based model achieving 93.66% accuracy in a centralized setting without privacy-preserving or explainability mechanisms. These limitations highlight the need for an integrated framework that combines federated learning, privacy preservation, explainability, and ensemble modelling.

In comparison, the proposed model achieves a competitive accuracy of 98.85% while simultaneously addressing multiple critical challenges. Unlike most existing studies, the proposed framework integrates federated learning, differential privacy, and explainability within a unified system. The use of a meta-ensemble stacking model combining Extra Trees, LightGBM, and CatBoost enhances predictive performance while maintaining robustness. Furthermore, the adoption of prediction-level aggregation simplifies federated learning and avoids model heterogeneity issues.

Importantly, the proposed framework provides all three key capabilities—federated learning, privacy preservation, and explainability—simultaneously, which is not achieved by any of the compared studies. While some models report slightly higher accuracy, they lack considerations for privacy, decentralization, or interpretability. Therefore, the proposed approach offers a more practical and clinically applicable solution by balancing performance with privacy and transparency, making it well-suited for real-world CKD prediction in distributed healthcare environments.

4 Discussion

The proposed privacy-preserving federated meta-ensemble stacking framework demonstrates strong performance in predicting chronic kidney disease (CKD) while addressing key challenges associated with clinical data, including heterogeneity, missing values, privacy concerns, and interpretability. The integration of multiple advanced components hybrid preprocessing, stacking-based ensemble learning, federated aggregation, differential privacy, and explainability enables the framework to achieve a balanced trade-off between predictive accuracy, robustness, and practical deployability.

The meta-ensemble stacking model plays a central role in enhancing predictive performance. By combining heterogeneous learners such as Extra Trees, LightGBM, and CatBoost through a meta-learner, the model effectively captures both linear and nonlinear relationships within the clinical dataset. The consistent performance observed across cross-validation folds indicates that the stacking approach successfully reduces both variance and bias, which are common limitations in single-model approaches. This is particularly important in medical datasets, where complex feature interactions often influence disease outcomes.

The robustness analysis further highlights the stability of the proposed model under realistic perturbations. Even with the introduction of Gaussian noise, the model exhibits only a marginal decline in performance, suggesting that it is resilient to measurement errors and variability inherent in clinical data collection. This robustness can be attributed to both the preprocessing strategy and the ensemble learning mechanism, which collectively enhance the model's ability to generalize beyond idealized conditions.

A key contribution of this study lies in the adoption of a federated learning paradigm tailored for healthcare applications. By distributing the dataset across multiple clients and performing decentralized training, the framework simulates real-world scenarios where data sharing is restricted due to privacy regulations. The strong cross-client generalization performance indicates that the model learns globally relevant patterns while adapting to local variations in data distributions. Unlike traditional federated approaches that rely on parameter aggregation, the use of prediction-level aggregation simplifies the learning process and mitigates issues related to model heterogeneity.

The incorporation of differential privacy further strengthens the framework by ensuring that sensitive patient information is protected during the learning process. The experimental results demonstrate that the addition of controlled Gaussian noise does not significantly compromise model performance, especially at moderate noise levels. This indicates an effective privacy–utility trade-off, which is critical for real-world deployment in healthcare systems where data confidentiality is paramount.

Another notable strength of the proposed framework is its interpretability through federated SHAP analysis. The identification of clinically significant features such as red blood cells, albumin, hypertension, and haemoglobin aligns well with established medical knowledge. Moreover, the consistency of feature importance across different clients reinforces the reliability of the model's decision-making process. The ability to provide both local and global explanations enhances transparency and builds trust among healthcare practitioners, which is essential for clinical adoption.

The communication efficiency analysis demonstrates that the framework is well-suited for deployment in resource-constrained environments. By incorporating model compression and sparse updates, the communication overhead is significantly reduced

without affecting predictive performance. This is helpful in distributed healthcare systems where resources like bandwidth and computing power are limited.

The external validation results further demonstrate the robustness of the proposed framework. While the benchmark UCI CKD dataset yielded the highest performance owing to its relatively compact feature space and well-curated clinical variables, the framework continued to achieve over 92% accuracy and 95% F1-score on two substantially larger independent datasets. The consistently high recall observed across all datasets indicates that the proposed model effectively identifies CKD cases while maintaining strong overall predictive performance. These findings confirm that the framework generalizes well across different patient populations and dataset characteristics, supporting its applicability to practical federated healthcare systems.

Although the experimental evaluation was performed using a benchmark CKD dataset containing 400 patient records, the proposed framework is designed to support substantially larger decentralized healthcare environments. In practical deployments, each participating institution independently trains its local model using its own patient records, while only privacy-preserving prediction information is exchanged with the aggregation server. As additional healthcare institutions participate, both the computational workload and storage requirements remain distributed across clients rather than centralized. Moreover, the communication optimization strategy reduces the volume of transmitted information, making the framework suitable for large-scale federated healthcare networks. Consequently, the proposed architecture is expected to scale effectively with increasing numbers of patients and participating institutions.

5 Strengths and limitations of the study

5.1 Strengths

The proposed framework presents several notable strengths that enhance its relevance for real-world clinical applications. First, the integration of a meta-ensemble stacking approach enables the model to effectively capture complex, non-linear relationships within clinical data. By combining heterogeneous learners such as Extra Trees, LightGBM, and CatBoost, the framework achieves improved generalization and reduces the risk of overfitting, which is particularly important in medical prediction tasks.

Second, the adoption of a federated learning paradigm addresses one of the most critical challenges in healthcare analytics—data privacy. By training models across clients without sharing raw data, the framework follows data protection rules and supports collaboration between institutions. The use of prediction-level aggregation further simplifies the federated setup and avoids issues related to model heterogeneity.

Another key strength lies in the incorporation of differential privacy mechanisms. The addition of controlled Gaussian noise ensures that sensitive patient information is protected while maintaining high predictive performance. This demonstrates a practical balance between privacy preservation and model utility, making the framework suitable for deployment in sensitive healthcare environments.

The inclusion of federated SHAP-based explainability significantly enhances the interpretability of the model. By providing both local and global feature importance across clients, the framework ensures transparency in decision-making. The consistency of identified features with established clinical indicators of CKD further validates the reliability of the model.

Additionally, the framework demonstrates strong robustness under data perturbations, indicating its ability to handle real-world variability in clinical measurements. The integration of communication-efficient strategies, such as model compression and sparse updates, further strengthens its applicability in resource-constrained environments, where bandwidth and computational limitations are common.

5.2 Limitations

Despite its strengths, the study has certain limitations that should be considered. One of the primary limitations is the relatively moderate size of the dataset used for evaluation. While the model demonstrates strong performance, smaller datasets may lead to optimistic results and limit the generalizability of findings. Future studies should validate the framework on larger and more diverse multi-institutional datasets.

Another limitation is the binary classification setup, which may not fully capture the complexity of CKD progression. In real-world clinical settings, CKD is often categorized into multiple stages, and extending the framework to multi-class classification could provide more clinically actionable insights.

The current implementation of federated learning relies on prediction-level aggregation rather than parameter sharing. While this simplifies the architecture and avoids heterogeneity issues, it may limit the ability to leverage deeper collaborative learning across clients. Exploring hybrid aggregation strategies could further enhance performance.

Finally, while SHAP-based explainability provides valuable insights, it may introduce additional computational overhead, especially in large-scale federated environments. Optimizing explainability methods for efficiency remains an important area for future research.

6 Feasibility analysis

The practical feasibility of the proposed privacy-preserving federated meta-ensemble framework was evaluated by considering its computational requirements, communication efficiency, scalability, privacy preservation, and interpretability. All experiments were successfully conducted on a standard computing platform equipped with an Intel Core i7 processor and 24 GB RAM without requiring GPU acceleration, demonstrating that the framework can be deployed using commonly available hardware.

From a communication perspective, the incorporation of compression and sparse update mechanisms reduced the communication cost from 0.82 MB to 0.25 MB, making the framework suitable for decentralized healthcare environments where network bandwidth may be limited. The prediction-level federated aggregation strategy further reduces communication complexity compared with conventional parameter-sharing approaches while maintaining high predictive performance.

The framework also satisfies important clinical deployment requirements by preserving patient privacy through differential privacy and enabling transparent decision-making using federated SHAP explanations. These characteristics support compliance with privacy regulations while improving clinician trust in model predictions.

Furthermore, the proposed framework achieved a mean accuracy of 98.85%, robust cross-client generalization of up to 99.25%, and stable performance under noisy conditions, indicating reliable operation across heterogeneous clinical environments. Collectively, these findings demonstrate that the proposed framework is computationally

efficient, communication-aware, privacy-preserving, and practically feasible for deployment in distributed healthcare systems supporting chronic kidney disease prediction.

7 Conclusion and future work

This study presents a prediction-level federated meta-ensemble stacking framework for CKD prediction that addresses challenges related to missing data, data variability, privacy preservation, and model interpretability. By integrating hybrid imputation, mutual information-based feature selection, a stacking-based ensemble model, prediction-level federated aggregation, differential privacy, and SHAP-based explainability, the proposed framework provides a unified and privacy-aware solution for reliable CKD prediction.

The experimental results demonstrate that the proposed framework achieves consistently high predictive performance while maintaining robustness under data perturbations. The prediction-level federated aggregation strategy enables effective knowledge sharing across decentralized healthcare institutions without exposing raw patient data, while ensuring strong generalization across clients. Furthermore, external validation on independent CKD datasets demonstrates the ability of the proposed framework to generalize beyond the original benchmark dataset. The integration of differential privacy further enhances the framework by protecting sensitive information while maintaining a favourable balance between data confidentiality and predictive performance.

Beyond predictive accuracy, the proposed framework emphasizes interpretability through federated SHAP analysis. The identification of clinically relevant features that remain consistent across multiple clients reinforces the reliability of the learned model and enhances its suitability for clinical decision support. In addition, the communication-efficient prediction-level aggregation strategy significantly reduces communication overhead, making the framework scalable and practical for deployment in resource-constrained healthcare environments.

Despite these promising outcomes, several avenues for future research remain. First, the framework can be evaluated on larger multi-centre datasets collected from multiple healthcare institutions to further validate its scalability and real-world generalizability. Incorporating longitudinal patient data may improve early CKD detection and disease progression modelling, while extending the current binary classification to multi-class prediction could better represent different CKD stages. Future work will also investigate fold-wise preprocessing within the cross-validation and federated training pipeline, where imputation, feature scaling, and feature selection are performed exclusively on local training data before being applied to validation or test data, thereby further strengthening methodological rigor and eliminating potential data leakage.

From a methodological perspective, future work will investigate advanced federated optimization strategies and stronger privacy-preserving mechanisms, including secure multi-party computation and homomorphic encryption, to provide stronger formal privacy guarantees. Privacy-preserving explainability techniques, including differentially private SHAP aggregation, also warrant further investigation. In addition, future studies will incorporate repeated experimental runs, confidence interval estimation, calibration analysis, and expected calibration error to provide a more comprehensive statistical evaluation of model reliability.

Practical deployment also requires addressing challenges related to heterogeneous hospital infrastructures, interoperability of electronic health record systems, compliance

with healthcare data privacy regulations (e.g., HIPAA and GDPR), secure communication protocols, and seamless integration with existing clinical workflows. Addressing these challenges will facilitate the adoption of the proposed framework in real-world healthcare settings.

In conclusion, the proposed prediction-level federated aggregation framework, combined with meta-ensemble learning, differential privacy, explainable AI, and communication-efficient aggregation, provides an accurate, robust, interpretable, and privacy-preserving solution for CKD prediction.

Acknowledgements
Not applicable.

Author contributions
KKN: Writing—review and editing, Supervision, DS, MVS: Methodology, Formal analysis. RP, CK: Investigation, Formal analysis. HLA, MSR: Review and editing, Conceptualization.

Funding
Not applicable.

Data availability
The dataset utilized in this study can be accessed at https://www.kaggle.com/datasets/mansoordaku/ckdisease. All code and materials used in this study are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Generative AI
Generative AI and AI-assisted technologies used only for formatting the content, grammar checking to improve the readability and language of the manuscript.

Author details
[1]Department of Artificial Intelligence and Data Science, Saveetha Engineering College, Chennai, Tamil Nadu, India
[2]Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, SIMATS, Chennai, India
[3]Department of Computer Science and Engineering (Artificial Intelligence), Madanapalle Institute of Technology & Science (MITS), Deemed to be University, Madanapalle, Andhra Pradesh, India
[4]School of Engineering and Applied Sciences, Kampala International University, Box 20000, Kampala, Uganda
[5]Department of Electronics and Communication Engineering, St Xavier's Catholic College of Engineering, Nagercoil, Tamil Nadu, India
[6]Department of Artificial Intelligence and Data Science, R.M.K. College of Engineering and Technology, Thiruvallur, Tamil Nadu, India
[7]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India

Received: 5 May 2026 / Accepted: 26 August 2026
Published online: 03 September 2026

References

1. Stewart S, Kalra PA, Blakeman T, Kontopantelis E, Cranmer-Gordon H, Sinha S. Chronic kidney disease: detect, diagnose, disclose—a UK primary care perspective of barriers and enablers to effective kidney care. BMC Med. 2024;22(1):331. https://doi.org/10.1186/s12916-024-03555-0.
2. Gogoi P, Valan JA. Machine learning approaches for predicting and diagnosing chronic kidney disease: current trends, challenges, solutions, and future directions. Int Urol Nephrol. 2025;57(4):1245–68. https://doi.org/10.1007/s11255-024-04281-5.
3. Delrue C, De Bruyne S, Speeckaert MM. Application of machine learning in chronic kidney disease: current status and future prospects. Biomedicines (Basel). 2024;12(3):568. https://doi.org/10.3390/biomedicines12030568.
4. Vashishth TK, Sharma V, Sharma KK, Kumar B, Chaudhary S, Panwar R. Securing the cloud: strategies for protecting sensitive patient data in cloud-based healthcare recommender systems. In: Healthcare recommender systems: techniques and recent developments, 2025. pp 185–220. https://doi.org/10.1007/978-3-031-80056-6_9.

5. Rani S, Kumar R, Panda BS, Kumar R, Muften NF, Abass MA, et al. Machine learning-powered smart healthcare systems in the era of big data: applications, diagnostic insights, challenges, and ethical implications. Diagnostics (Basel). 2025;15(15):1914. https://doi.org/10.3390/diagnostics15151914.
6. Ji S, Tan Y, Saravirta T, Yang Z, Liu Y, Vasankari L, et al. Emerging trends in federated learning: from model fusion to federated x learning. Int J Mach Learn Cybern. 2024;15(9):3769–90. https://doi.org/10.1007/s13042-024-02119-1.
7. Khraisat A, Talukder MA, Uddin MA, Alazab A. RF-FedAvg: federated learning-based random forest model for intrusion detection in wireless sensor networks. Cluster Comput. 2025;28(13):873. https://doi.org/10.1007/s10586-025-05591-8.
8. Talukder MA, Khalid M, Sultana N. A hybrid machine learning model for intrusion detection in wireless sensor networks leveraging data balancing and dimensionality reduction. Sci Rep. 2025;15(1):4617. https://doi.org/10.1038/s41598-025-87028-1.
9. Talukder MA, Sharmin S, Uddin MA, Islam MM, Aryal S. MLSTL-WSN: machine learning-based intrusion detection using SMOTETomek in WSNs. Int J Inf Secur. 2024;23(3):2139–58. https://doi.org/10.1007/s10207-024-00833-z.
10. Albshaier L, Almarri S, Albuali A. Federated learning for cloud and edge security: a systematic review of challenges and AI opportunities. Electronics (Basel). 2025;14(5):1019. https://doi.org/10.3390/electronics14051019.
11. Tariq A, Serhani MA, Sallabi FM, Barka ES, Qayyum T, Khater HM, et al. Trustworthy federated learning: a comprehensive review, architecture, key challenges, and future research prospects. IEEE Open J Commun Soc. 2024;5:4920–98. https://doi.org/10.1109/OJCOMS.2024.3438264.
12. Liu Y, Liu C, Zheng J, Xu C, Wang D. Improving explainability and integrability of medical AI to promote health care professional acceptance and use: mixed systematic review. J Med Internet Res. 2025;27:e73374. https://doi.org/10.2196/73374.
13. Saifullah S, Mercier D, Lucieri A, Dengel A, Ahmed S. The privacy-explainability trade-off: unraveling the impacts of differential privacy and federated learning on attribution methods. Front Artif Intell. 2024;7:1236947. https://doi.org/10.3389/frai.2024.1236947.
14. Mahmoud AS, Lamouchi O, Belghith S. Advancements in machine learning and deep learning for early diagnosis of chronic kidney diseases: a comprehensive review. Babylon J Mach Learn. 2024. https://doi.org/10.58496/BJML/2024/015.
15. Guido R, Ferrisi S, Lofaro D, Conforti D. An overview on the advancements of support vector machine models in healthcare applications: a review. Information (Basel). 2024;15(4):235. https://doi.org/10.3390/info15040235.
16. Iliyas II, Boukari S, Gital AYU. Recent trends in prediction of chronic kidney disease using different learning approaches: a systematic literature review. J Med Artif Intell. 2025;8:62. https://doi.org/10.21037/jmai-24-256.
17 Yuan S, Guo L, Xu F. Artificial intelligence in nephrology: predicting CKD progression and personalizing treatment. Int Urol Nephrol. 2025;58(7):2615–45. https://doi.org/10.1007/s11255-025-04878-4.
18. Mohammed FA, Tune KK, Assefa BG, Jett M, Muhie S. Medical image classifications using convolutional neural networks: a survey of current methods and statistical modeling of the literature. Mach Learn Knowl Extr. 2024;6(1):699–735. https://doi.org/10.3390/make6010033.
19. Dangi RR, Sharma A, Vageriya V. Transforming healthcare in low-resource settings with artificial intelligence: recent developments and outcomes. Public Health Nurs. 2025;42(2):1017–30. https://doi.org/10.1111/phn.13500.
20 Hassija V, Chamola V, Mahapatra A, Singal A, Goel D, Huang K, et al. Interpreting black-box models: a review on explainable artificial intelligence. Cognit Comput. 2024;16(1):45–74. https://doi.org/10.1007/s12559-023-10179-8.
21. Li N, Lewin A, Ning S, Waito M, Zeller MP, Tinmouth A, et al. Privacy-preserving federated data access and federated learning: improved data sharing and AI model development in transfusion medicine. Transfusion. 2025;65(1):22–8. https://doi.org/10.1111/trf.18077.
22. Abbas SR, Abbas Z, Zahir A, Lee SW. Federated learning in smart healthcare: a comprehensive review on privacy, security, and predictive analytics with IoT integration. Healthcare (Basel). 2024;12(24):2587. https://doi.org/10.3390/healthcare12242587.
23. Lu Z, Pan H, Dai Y, Si X, Zhang Y. Federated learning with non-iid data: a survey. IEEE Internet Things J. 2024;11(11):19188–209. https://doi.org/10.1109/JIOT.2024.3376548.
24. Yang Z, Yan X, Chen G, Tian X. Mitigating membership inference attacks via generative denoising mechanisms. Mathematics. 2025;13(19):3070. https://doi.org/10.3390/math13193070.
25. Thantharate P, Bhojwani S, Thantharate A. DPShield: optimizing differential privacy for high-utility data analysis in sensitive domains. Electronics. 2024;13(12):2333. https://doi.org/10.3390/electronics13122333.
26. Vimbi V, Shaffi N, Mahmud M. Interpreting artificial intelligence models: a systematic review on the application of LIME and SHAP in Alzheimer's disease detection. Brain Inform. 2024;11(1):10. https://doi.org/10.1186/s40708-024-00222-1.
27 Aristeidis K, Anastasios G, Natalia A, Christos K. Next-gen explainable AI (XAI) for federated and distributed internet of things systems: a state-of-the-art survey. Future Internet. 2026;18(2):83. https://doi.org/10.3390/fi18020083.
28. Islam MA, Majumder MZH, Hussein MA. Chronic kidney disease prediction based on machine learning algorithms J. Pathol Inform. 2023;14:100189. https://doi.org/10.1016/j.jpi.2023.100189.
29. Khafaga DS, Khodadadi N, Khodadadi E, Ali Alhussan A, Eid MM, El-Kenawy ESM. Enhanced early chronic kidney disease prediction using hybrid waterwheel plant algorithm for deep neural network optimization. Sci Rep. 2025. https://doi.org/10.1038/s41598-025-26382-6.
30. Rehman MU, Naseem S, Butt AUR, Mahmood T, Khan AR, Khan I, et al. Predicting coronary heart disease with advanced machine learning classifiers for improved cardiovascular risk assessment. Sci Rep. 2025;15(1):13361. https://doi.org/10.1038/s41598-025-96437-1.
31. Dimlo UMF, Banoth S, Bihade P, Mjery Y, Kumar KJ, R U, et al. Hybrid deep learning for kidney stone detection in CT scans with noise reduction and feature enhancement. J Comput Sci. 2026;22(3):813–25. https://doi.org/10.3844/jcssp.2026.813.825.
32. Al Ghamdi M, Alyahyan S. Enhanced chronic kidney disease detection: a hybrid deep learning framework using clinical biomarkers and ensemble feature engineering with DeepCKD-Net. Appl Sci. 2026;16(6):3024. https://doi.org/10.3390/app16063024.
33 Ur A, Asif M, Qasim KU, Ishak MK, Khan I, Khan J, et al. AdvancedHybridNet: an AI-powered hybrid ensemble for high-accuracy thyroid disease diagnosis using dynamic feature selection. Endocr Syst Dis. 2025. https://doi.org/10.21203/rs.3.rs-7401588/v1.

34. Butt AUR, Hamid I, Nawaz Q, Mahmood T, Zhang X, Yaqub M. A novel multi-scale deep learning approach for the early detection of Alzheimer's disease using fMRI. In: Proceedings of the 2024 5th international conference on computer, big data and artificial intelligence (ICCBD+AI). IEEE; 2024. Pp 85–90. https://doi.org/10.1109/ICCBD-AI65562.2024.00022.
35. Butt AUR, Asif M, Rashid T, Ali H, Raoof S, Raoof F. Beyond boundaries: a novel ensemble approach for breast cancer detection in ultrasound imaging using deep learning. In: Proceedings of the 2024 21st international Bhurban conference on applied sciences and technology (IBCAST). IEEE; 2024. pp 315–320. https://doi.org/10.1109/IBCAST61650.2024.10877027.
36. Quan MK, Pathirana PN, Wijayasundara M, Setunge S, Nguyen DC, Brinton CG, et al. Federated learning for cyber physical systems: a comprehensive survey. IEEE Commun Surv Tutor. 2025. https://doi.org/10.1109/COMST.2025.3570288.
37 Komal KN, Tulasi RL, Vigneswari D. An ensemble multi-model technique for predicting chronic kidney disease. Int J Electr Comput Eng. 2019;9(2):1321. https://doi.org/10.11591/ijece.v9i2.pp1321-1326.
38. Jahin GTS, Maliha F, Hassan FB, Muztaba AMA, Patwary MJA, Arefin MS. Privacy-preserving and explainable multi-hospital CKD prediction via federated learning with Blockchain logging and differential privacy. In: 2025 IEEE 2nd international conference on computing, applications and systems *(COMPAS)*. IEEE; 2025. pp 1–6. https://doi.org/10.1109/COMPAS67506.2025.11381661.
39. Bijoy MHI, Mia MJ, Rahman MM, Arefin MS, Dhar PK, Shimamura T. A robot process automation based mobile application for early prediction of chronic kidney disease using machine learning. Discov Appl Sci. 2025;7(6):528. https://doi.org/10.1007/s42452-025-06980-9.
40. Disease K, Group IGOKCW. KDIGO 2024 clinical practice guideline for the evaluation and management of chronic kidney disease. Kidney Int. 2024;105(4S):S117–314. https://doi.org/10.1016/j.kint.2023.10.018.
41. Levey AS, Coresh J. Chronic kidney disease. Lancet. 2012;379(9811):165–80. https://doi.org/10.1016/S0140-6736(11)60178-5.
42. Pathak A, Singh OP, Biswas U. Predicting chronic kidney disease using machine learning methodologies. In: 2025 8th international conference on electronics, materials engineering and nano-technology (IEMENTech). IEEE; 2025. pp. 1–6. https://doi.org/10.1109/ICAISET66439.2026.11541350.
43. Chaudhry A. Chronic kidney disease synthetic dataset. Kaggle. 2025. https://www.kaggle.com/datasets/amna0chaudhry/chronic-kidney-disease-synthetic-dataset.
44. Fauzan FJ, Putri CM, Laura P. Performance comparison of classification algorithms for chronic kidney disease prediction. IJATIS Indonesian J Appl Technol Innov Sci. 2024;1(2):68–74. https://doi.org/10.57152/ijatis.v1i2.1120.
45. Ganie SM, Dutta Pramanik PK, Mallik S, Zhao Z. Chronic kidney disease prediction using boosting techniques based on clinical parameters. PLoS ONE. 2023;18(12):e0295234. https://doi.org/10.1371/journal.pone.0295234.
46. Dritsas E, Trigka M. Machine learning techniques for chronic kidney disease risk prediction. Big Data Cogn Comput. 2022;6(3):98. https://doi.org/10.3390/bdcc6030098.
47. Halder RK, Uddin MN, Uddin MA, Aryal S, Saha S, Hossen R, et al. ML-CKDP: machine learning-based chronic kidney disease prediction with smart web application. J Pathol Inform. 2024;15:100371. https://doi.org/10.1016/j.jpi.2024.100371.
48. Arif MS, Rehman AU, Asif D. Explainable machine learning model for chronic kidney disease prediction. Algorithms. 2024;17(10):443. https://doi.org/10.3390/a17100443.
49. Zuriati Z, Meilantika D, Arpan A, Permata R, Sriyanto S, Mas' ud MZ. Enhancing chronic kidney disease classification using decision tree and bootstrap aggregating: UCI dataset study with improved accuracy and AUC-ROC. JurnalTeknikInformatika (Jutif). 2025;6(5):3111–23. https://doi.org/10.52436/1.jutif.2025.6.5.5271.
50. Hegde MG, Ruthvika B, Jain RB, Shenoy PD, Venugopal KR, Canchi A. A privacy-preserving federated learning method with homomorphic encryption for chronic kidney disease stage prediction. Eng Technol Appl Sci Res. 2025;15(4):26019–26. https://doi.org/10.48084/etasr.11928.
51. Zhao Y, Lu S, Lu J, Yang L, Lo CW, Wong MK, et al. Risk prediction of chronic kidney disease progression in type 2 diabetes mellitus across diverse populations. npj Digit Med. 2026;9(1):250. https://doi.org/10.1038/s41746-026-02439-2.
52. Dedgaonkar SG, Navale GS, Shelke P, Dhumane AV. Chronic kidney disease detection using XceptionNet with harmonic addax optimization. Comput Biol Chem. 2026. https://doi.org/10.1016/j.compbiolchem.2026.108921.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.