SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V13I6P116

Volume 13 Issue 6, 195-206, June 2026
© 2026 Seventh Sense Research Group®

Original Article

A Hybrid Multimodal Framework for Accurate CKD Stage Prediction via Ensemble Learning and DenseNet

Utkarsha Rane[1], Satish Ramesh Kolhe[2]

[1,2]*School of Computer Sciences, Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon, India.*

[1]*Corresponding Author : utkarsharane05@gmail.com*

Received: 18 March 2026 Revised: 17 April 2026 Accepted: 16 May 2026 Published: 27 June 2026

Abstract *- The gradual and irreversible condition known as Chronic Kidney Disease (CKD) that needs precise stage-wise prediction for early clinical intervention. The traditional diagnostic methods are primarily based on biochemical parameters, which may not effectively address the structural changes in the kidney status. This paper presents a multimodal machine learning framework that combines structured clinical information and kidney imaging for better chronic kidney disease stage prediction. Two different datasets were used: a hospital-retrieved tabular dataset of 500 patients with demographic and biochemical parameters, and an accessible public computed tomography dataset of 12,446 kidney pictures in the collection. Ensemble machine learning algorithms, such as Light Gradient Boosting Machine, Random Forest, and Extreme Gradient Boosting, were used for clinical dataset analysis, and a DenseNet-based convolutional neural network was used for image analysis. The Light Gradient Boosting Machine performed the best on the tabular data with an accuracy of 92.43% and a region beneath the curve of 0.988. The DenseNet model performed with 99.28% accuracy and high precision and recall values for all classes of images. The multimodal fusion approach combined the independent predictions of both datasets to improve the reliability of classification at the intermediate level. The findings show that the use of clinical and imaging modalities with separate datasets improves the accuracy of prediction and provides a reliable decision support method for evaluating chronic renal disease stages.*

Keywords **-** *Chronic Kidney Disease, Multimodal Learning, Gradient Boosting, Densenet, Medical Decision Support.*

1. Introduction

CKD, or Chronic renal disease, has emerged as a significant global health concern, affecting a rapidly growing population worldwide. One of the key challenges associated with CKD is its silent progression, where symptoms often become apparent only at advanced stages, limiting opportunities for timely intervention [1]. Consequently, accurate early diagnosis and precise staging are critical for slowing disease progression, reducing complications such as cardiovascular events, and improving long-term patient outcomes. Clinical guidelines emphasize the importance of stage-specific diagnosis, typically relying on biochemical markers like serum creatinine and Estimated Glomerular Filtration Rate (eGFR), and albuminuria. Although these indicators form the foundation of CKD assessment, their effectiveness in real-world clinical settings is often constrained by inter-patient variability, comorbid conditions, and biological heterogeneity [2].

Over the past few years, the increasing availability of structured clinical datasets and biomedical imaging has led to the growing adoption of Machine Learning (ML) and Artificial Intelligence (AI) techniques in healthcare decision support systems. ML models trained on clinical data can uncover complex nonlinear relationships that are not easily captured by traditional statistical methods. Algorithms such as Random Forests, Support Vector Machines, and Gradient Boosting Machines have shown strong performance in disease classification tasks based on laboratory and demographic data [3]. Furthermore, ensemble learning strategies have improved prediction stability and reduced overfitting by combining multiple models. However, despite these advancements, models relying solely on tabular data remain limited in their ability to fully capture the multifaceted nature of CKD progression, which involves both physiological and structural changes [4].

Parallel developments in medical imaging have further enhanced diagnostic capabilities, particularly through the use of ultrasound imaging, which is widely regarded as a non-invasive and cost-effective modality for renal assessment. The application of deep learning, especially Convolutional Neural Networks (CNNs), has significantly advanced medical image analysis by enabling automated feature extraction and identification of subtle anatomical patterns [5]. Architectures such as DenseNet, ResNet, and VGG have demonstrated strong performance in tasks including lesion detection, segmentation, and organ classification. In the context of renal imaging, CNNs can effectively learn structural features such as kidney morphology, texture of the kidney parenchyma, and presence of abnormalities such as cysts and stones. Nevertheless, it is not enough to rely solely on imaging methods, as structural changes do not necessarily correlate with biochemical changes [6].

Despite improvements in both tabular and imaging-based methods, there is still a considerable research gap. The existing research has mostly concentrated on single modality data, either clinical or imaging. The drawback is that they lack diagnostic comprehensiveness.

The models based on clinical data alone do not account for anatomical changes, and models based on imaging data alone do not account for biochemical changes. There is little research done on multi-class prediction for CKD stages using integrated data from multiple modalities, especially with strong validation for various stages of disease. The ambiguity is highest in intermediate stages.

To overcome these limitations, Recent studies have concentrated on creating frameworks for multimodal learning that can successfully combine diverse input sources, including clinical, demographic, and imaging information. These models seek to offer a more holistic perspective on the patient's health by incorporating physiological and morphological information. Indeed, multimodal machine learning has been successful in enhancing predictive accuracy and offering personalized decision-making. However, many multimodal models are ineffective in terms of fusion strategies and performance evaluation for all CKD stages.

In consideration of these challenges, this study aims to develop a unified multimodal framework for CKD stage prediction that utilizes both structured clinical information and imaging-based features. This is achieved by combining ensemble learning methods for structured data with deep learning for ultrasound imaging, thus allowing for simultaneous extraction of biochemical and structural information. This framework is developed to tackle the challenges associated with CKD staging, including class imbalance, subtle inter-stage differences, and overlapping clinical features.

The novelty in this research is based on the creation of a new architecture for multimodal learning, which is able to combine effectively both ensemble-based tabular models and deep CNN models for imaging analysis in multiclass prediction for CKD stages.

Unlike previous research in this area, which has only used single modality data and limited fusion methods, this research utilizes multiple data sources in an integrated prediction framework. Several performance criteria are also utilized to measure the model's efficacy, like recall, accuracy, precision, and F1-score for every stage of chronic kidney disease.

Moreover, the integration of biochemical, demographic, and imaging characteristics in the proposed framework ensures a more comprehensive solution for the staging of CKD in a clinically relevant manner. The experimental results clearly indicate the advantage of multimodal fusion in the accuracy of the diagnosis and the ability to differentiate the stages, even in the more challenging stages of the disease. This would be beneficial to the creation of intelligent systems that could aid in the diagnosis and intervention at the early stages of the disease.

This paper's remaining sections are organized as follows: The review of relevant works is in Section 2 on CKD prediction and multimodal learning. Section 3 introduces the methodology used in this research, including data preparation, feature engineering, and model design. Section 4 reports the experimental results and comparison. The study is discussed in Section 5, and Section 6 is a conclusion of this research, including suggestions for future research.

2. Related Work

The latest developments in the forecasting of CKD can be broadly classified into three categories: (i) applying models for machine learning using structured clinical data, (ii) leveraging medical data to apply deep learning algorithms to imaging data, and (iii) the application of emerging multimodal frameworks using diverse data sources. Although each of the above approaches has shown promising results in the area of predicting CKD, the use of each of the above approaches in isolation may not be sufficient.

However, a considerable amount of research has been carried out on using traditional and hybrid approaches, and algorithms used in machine learning. In this context, a comparative analysis of traditional and ensemble-based approaches using SVM, KNN, and LightGBM was conducted by Islam [8]. In this research, it has been shown that tree-based ensemble approaches are more effective for CKD classification. Another research by Subashini [9] has been conducted using feature selection techniques and ensemble-based approaches to improve the stability of prediction and class imbalance. However, this research has mainly concentrated on non-image-based approaches. Another research by Halder [10] has been conducted using a new framework called ML-CKDP, which mainly focuses on preprocessing methods such as balancing and imputation to enhance the generalization of the approaches. However, these approaches are mainly restricted to using tabular data and cannot be used to extract structural information related to the kidneys.

Deep learning models have also been used extensively in medical imaging for renal analysis. Miguel [11] demonstrated the efficacy of CNN-based models in ultrasound image analysis. In this work, Miguel used data augmentation to obtain promising results. However, this work is focused on prenatal anomalies rather than CKD staging. Alsekait [12] introduced ensemble deep learning models that incorporate feature selection. This work improved sensitivity in CKD detection. Maqsood [13] introduced hybrid models of CNNs for CT image classification in renal pathology. This work demonstrated the advantages of using hybrid models. Although this work demonstrated excellent results in medical imaging, it is difficult to interpret. Chen [14] also demonstrated the efficacy of ensemble learning models. This work demonstrated excellent accuracy in CKD detection.

Surveys conducted in recent times have emphasized the importance of multimodal AI in the healthcare sector. Simon in [15] presented a comprehensive review of the integrated imaging and clinical models, with special emphasis on the translational challenge. Liu in [16] presented various data fusion strategies, such as early, late, and attention-based fusion strategies. This offers a fundamental foundation for the growth of the multimodal system.

Recent studies also emphasize the various aspects of CKD-related AI research. Boosting-based techniques have been successful in improving predictive accuracy. However, they are also highly susceptible to noise [17]. Multimodal CNN-based fusion techniques have also been explored to help enhance the early identification of fibrosis. These techniques are also limited to specific tasks [18]. Recent advancements in segmentation techniques using dual attention models have been successful in improving the analysis of kidney structures.

However, this is not related to CKD stage classification [19]. Ensemble learning using CNN models has also been successful in improving accuracy even in noisy imaging conditions. These techniques are also limited to single-mode learning [20]. Techniques such as stacking ensemble learning have been successful in improving stability in regional dataset predictions.

However, this is limited to learning without imaging data [21]. Transfer learning-based techniques have been explored for improving CT imaging. However, this is limited to ultrasound-based CKD workflows [22]. Review studies also emphasize that ensemble learning is more effective than individual learning. CNN-based ultrasound analysis is also successful in specific renal conditions such as IgA nephropathy [24].

To overcome the shortcomings of the current methods, the present study proposes a multimodal framework for predicting the stages of CKD by combining the approaches of tabular learning with the help of the ensemble technique and DenseNet-based imaging analysis. Unlike the existing approaches, the present research proposes a paradigm that incorporates the biochemical, demographic, and structural information in a unified framework for the purpose of predicting the stages of CKD.

3. Methodology

The current study's objective is to create a holistic framework for multimodal learning in the prediction of CKD stages, where tabular information and imaging representation learning are integrated together. The original dataset used in the present work contains the following information related to the kidney: Gender, Serum Creatinine Sc, EGFR, ACR, UPCR, and BUN-Creatinine Ratio. The samples in the original dataset were balanced, as the stages of CKD were considered. Figure 1 presents the proposed methodological framework, where the major steps include holistic preprocessing, tabular modeling by applying ensemble learning, dense net-based imaging feature learning, multimodal fusion, and softmax-based classification. Each component of the methodological framework has the potential to ensure the development of robust prediction with interpretability for different cases of patients.

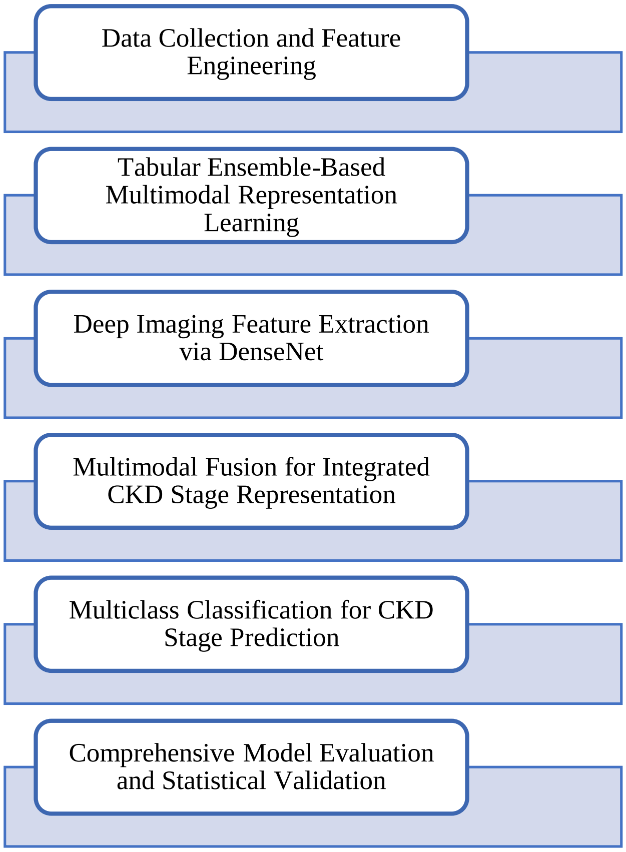

Fig. 1 Multimodal CKD stage prediction methodology framework

3.1. Data Acquisition and Structured Feature Preparation

Two different datasets were used in this research to develop a multimodal framework for CKD stage prediction. One is a clinical tabular dataset that was obtained from a hospital, and it contains information on demographic and biochemical factors such as age, gender, creatinine, eGFR, ACR, UPCR, BUN/creatinine, hemoglobin, blood pressure, and race. From this dataset, information on a total of 500 patients is available, ranging in age from 25 to 84 years, with a mean of 55.39. There are 252 males and 248 females in this dataset. The CKD stage, which is a dependent variable, varies from 1 to 5, with a mean of 2.4. This indicates that most of the people in this dataset are in stage 2 or 3 of CKD. There is a high prevalence of hypertension in this dataset, at a rate of 65.6%, and diabetes, at a rate of 53%.

The second dataset used in the study was a publicly available dataset with renal CT scans. The collection was titled "CT Kidney Dataset: Normal–Cyst–Tumor–and–Stone" [25], available at Kaggle at https://www.kaggle.com/datasets/nazmul0087/ct-kidney-dataset-normal-cyst-tumor-and-stone/data. The dataset consists of 12,446 anonymized images collected from the PACS system of different hospitals.

All the images have been confirmed to be of diagnostic quality. The images have been classified into four types: normal, cyst, stone, and tumor, with 5,077, 3,709, 1,377, and 2,283 images, respectively. All the images are in lossless formats like JPG. The images are anonymized, with no information available regarding the patients.

The preprocessing techniques for the tabular data included cleaning, normalizing numerical features, encoding categorical features, class balancing, and splitting the data. These two sets of data, when used together, offer a comprehensive multimodal approach by combining both image and clinical data, thus improving the predictive accuracy of CKD stages.

3.1.1. Data Cleaning and Harmonization

The structured clinical dataset consists of 500 patients with demographic information, clinical laboratory information, and comorbid conditions for the prediction of CKD stages (Stage 1-5). In the data cleaning process, the data is checked for completeness with respect to all the major renal-related information, such as Serum Creatinine, eGFR, BUN, ACR, UPCR, Hemoglobin, Blood Pressure, Albumin, and Blood Glucose.

In the process of handling missing data, the data is filled using appropriate imputations where necessary. In this process, no data of any patient should be removed. In the process of identifying outliers in the data, statistical range analysis is conducted on the data.

This process is mainly carried out on highly varying data such as Serum Creatinine (0.61-7.96) and UPCR (30-482.1). Continuous variables were normalized through standard scaling to reduce variability and skewness. Categorical variables such as Gender, Diabetes, and Hypertension were converted to numerical variables for consistency. This will ensure that the structural compatibility in the imaging pipeline is maintained for multimodal integration.

3.1.2. Feature Engineering and Diagnostic Parameter Construction

Feature engineering was based on improving both clinical interpretability and prediction strength. Calculated metrics like the BUN: Creatinine ratio were generated to better approximate renal filtration capacity. The combination of ACR and UPCR usage was retained to retain information on proteinuria. As eGFR has a mean of 68.10 and CKD stage has a mean of 2.4, correlation analysis was performed to avoid redundancy. Continuous variables are retained to retain precision.

3.1.3. Dataset Balancing and Train–Test Structuring

The dataset was divided using the approach of stratified sampling, where the ratio is 80:20. The SMOTE technique was applied only on the training set to prevent stage imbalance. Five-fold cross-validation also ensured an additional layer of stability in the model, while the test set remained completely unseen.

3.2. Multimodal Representation Modeling Through Ensemble Tabular Learning

The ensemble machine learning algorithms are employed to learn the nonlinear correlations between the structured clinical dataset of 500 patients, which includes demographic factors (Age, Gender), laboratory biomarkers (Serum Creatinine, BUN, eGFR, ACR, UPCR, Hemoglobin, Albumin, Blood Glucose, Blood Pressure), and their respective comorbidities (Diabetes, Hypertension). These factors, taken together, reflect the biochemical and physiological manifestations of disease progression at each Stage 1-5 of CKD.

Gradient boosting, random forests, and stacking techniques were applied to learn the complex interactions between the renal biomarkers, where several biomarkers are known to have large variability, including Serum Creatinine (0.61-7.96) and UPCR (30-482.1). The resulting tabular models output probability vectors for CKD stage, which are then combined with the imaging features extracted using DenseNet. The structured representation of biochemical data allows for increased sensitivity to the biochemical severity while complementing structural imaging features.

3.2.1. Gradient-Boosted Feature Learning

This submethod involves the use of gradient boosting for the sequential reduction of residual errors to refine predictive performance:

$$F_m(x) = F_{m-1}(x) + \eta h_m(x) \tag{1}$$

Where, x = input feature vector (clinical features), $F_m(x)$ = ensemble prediction at iteration m, $F_{m-1}(x)$ = prediction from previous iteration, η = learning rate (shrinkage factor), $h_m(x)$ = weak learner (decision tree) trained on residuals, and m = boosting iteration index

Herein, x is the clinical feature vector of every patient, comprising scaled laboratory and demographic variables. The boosting process will iteratively enhance the classification sensitivity to subtle differences in renal function, especially across adjacent stages, for instance, Stage 2–3, thereby strengthening the multiclass discrimination.

3.2.2. Random Forest Clinical Pattern Extraction

Random Forests were used to reduce variance and improve the robustness via bootstrap aggregation:

$$\hat{y} = \mathrm{mode}(h_1(x), h_2(x), \ldots, h_T(x)) \tag{2}$$

Where, $\hat{y}$ = final predicted class, x = input feature vector, $h_t(x)$ = prediction of the t-th decision tree, T = total number of trees, and mode(.) = majority voting operator.

This approach profiles various renal signatures in patients with a high prevalence of hypertension 65.6% and diabetes 53%. It reduces overfitting in noisy laboratory measurements by aggregation across many trees and identifies the most stable predictive biomarkers.

3.2.3. Stacked Ensemble Probability Encoding

The work combines outputs of multiple ensemble learners using a meta-model. The combination is expressed as:

$$P_{final} = g(P_1, P_2, \ldots P_n) \tag{3}$$

Where P_{final} = final combined probability vector, and g is a trained meta-learner.

Stacking is a method that incorporates complementary learner strengths into unified CKD stage probability scores. These probability vectors are then used as structured inputs for multimodal fusion with imaging features, thereby improving the overall performance of stage prediction.

3.3. Imaging Feature Modeling Using DenseNet Convolutional Architecture

This study complements the structured clinical data set of 500 CKD patients with a DenseNet-based deep learning pipeline for feature extraction from 12,446 renal CT images, which are classified into normal, cyst, stone, and tumor. Although the structured data set predicts CKD Stage (1-5), the proposed imaging pipeline extracts the structural and morphological variations of the kidneys, which are indicative of the severity of the illness. The pictures are normalized and resized. to ensure consistency in computation. DenseNet's densely connected architecture enables efficient gradient flow, which helps in learning deeper features without any redundancy. The model learns the high-level imaging features, which are then combined with the probability vectors learned from the ensemble of tabular models. This study closes the gap between learning ideas from structural imaging features as well as biochemical markers like eGFR, Serum Creatinine, and ACR.

3.3.1. Dense Connectivity Feature Flow

This makes use of the dense connectivity rule of DenseNet:

$$x_\ell = H_\ell([x_0, x_1, \ldots, x_{\ell-1}]) \tag{4}$$

Where, x_l = output feature map of layer l, H_l (·) = composite transformation (BatchNorm + ReLU + Conv), [·] = concatenation operator, and $x_0,\ldots,x_{l-1}$ = outputs from all preceding layers.

The dense connections facilitate feature reuse across layers. This helps the model to capture even subtle features of renal textures. This is particularly important when capturing structural variations that reflect subtle differences in CKD severity levels. The architecture reduces vanishing gradients and improves multiscale representation learning.

3.3.2. Convolutional Renal Embedding Extraction

This work uses convolution operations based on:

$$f(x) = W * x + b \tag{5}$$

Where, x = input feature map, W = convolution kernel (filter weights), * = convolution operator, b = bias term, and $f(x)$ = output feature map

Convolutions enable the derivation of spatial and textural renal descriptors such as density variations and structural irregularities. Such an imaging embedding is of high dimensionality.

3.3.3. DenseNet Feature Aggregation and Global Pooling

This study uses global average pooling to compress the deep features to compact embedding vectors. As such, the pooling operation is defined as:

$$z_k = \frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W} x_{k,i,j} \tag{6}$$

Where, z_k = pooled value of feature map k, H, W = height and width of feature map, $x_{k,i,j}$ = activation at position (i, j) of channel k.

This dimensionality reduction maintains essential information for diagnosis, following fusion with a tabular form of probability vectors for multimodal CKD stage prediction.

3.4. Multimodal Fusion Layer for Unified CKD Stage Representation

This final stage of CKD prediction involves the combined output of both modalities: (i) the probability vectors obtained from the ensemble models trained on the 500-patient structured clinical dataset, and (ii) the high-dimensional imaging embeddings obtained from the 12,446 CT scans using DenseNet. The tabular branch corresponds to the indicators of severity quantified via the blood test and demographic factors of patients. This includes values for eGFR (mean 68.10), Serum Creatinine (0.61 to 7.96), ACR, UPCR, and the prevalence of comorbidity of hypertension (65.6%) and diabetes (53%), while the imaging branch corresponds to the structural differences in the kidney morphology.

3.4.1. Hybrid Concatenation-Based Fusion

This work aligns the tabular probability vector Pt and imaging embedding E_i.

$$F = [P_t \parallel E_i] \tag{7}$$

Where, F = fused feature vector, P_t = tabular probability vector from ensemble, E_i = imaging embedding from DenseNet, and ∥ = concatenation operator.

The F vector is used for the fused feature vector. The concatenation process maintains the complete distribution of the probabilistic stage. This vector incorporates synergistic interaction between laboratory abnormalities and anatomical features into a multiclass CKD stage representation.

3.4.2. Fully Connected Projection Transformation

This project fused features into a compact latent space

$$h = \sigma(WF + b) \tag{8}$$

Where, F = fused feature vector, W = weight matrix, b = bias vector, $\sigma(\cdot)$ = nonlinear activation function (e.g., ReLU), and h = projected latent representation.

This transformation improves the separability among CKD Stage 1–5 by learning nonlinear interactions between biochemical severity markers and imaging features. The projection emphasizes salient multimodal characteristics while reducing redundancy.

3.4.3. Regularization and Dropout Stabilization

Considering that the sample size of tabular data was moderate (500 patients), L2 regularization and dropout were used in fully connected layers to avoid overfitting. These mechanisms improve generalization across heterogeneous patient profiles and ensure stable performance during final CKD stage classification.

3.5. Multiclass Classification Architecture for CKD Stage Prediction

The joint model for multiple modalities, where there is an integration of tabular probability picked from the 500-patient population and DenseNet model learned image representations from 12,446 CT images, is then went through a softmax activation and a completely connected classification head function. The final model is designed to predict CKD Stage (1-5) by incorporating biochemical markers such as eGFR, Serum Creatinine, ACR, UPCR, BUN, and structural markers from the kidney. The model parameters were learned through stochastic gradient-based learning with categorical cross-entropy, enabling the boundary to learn while keeping it probabilistically interpretable.

3.5.1. Softmax Stage Probability Computation

This computes class likelihood using:

$$p(y = k \mid x) = \frac{e^{z_k}}{\sum_{j=1}^{6} e^{z_j}} \tag{9}$$

Where, $p(y = k \mid x)$ = probability of class k, z_k = logit (pre-softmax score) for class k, j = class index, and 6 = total number of CKD stages.

Softmax normalizes the probability outputs across the stages of CKD, giving interpretable confidence scores to each stage. This is a probabilistic formulation that allows clinical transparency in the outcomes of predictions.

3.5.2. Cross-Entropy Loss Optimization

This research minimizes the objective:

$$L = -\sum_{k=1}^{6} y_k \, log(P_k) \tag{10}$$

Where, L = cross-entropy loss, y_k = ground truth indicator (1 if class k, else 0), P_k = predicted probability for class k, and k = class index

Moreover, this loss will also penalize incorrect stage assignments as well as the separation between adjacent CKD stages, e.g., Stage 2-3, which show partially overlapping biochemical profiles.

3.5.3. Adaptive Parameter Tuning

Additionally, learning rate scheduling, gradient clipping, and weight decay were used to stabilize the convergence for the multimodal framework. These strategies improve generalization performance. They ensure stable optimization for heterogeneous clinical and imaging representations.

3.6. Model Evaluation, Statistical Validation, and Performance Assessment

In this research work, the efficiency of the recommended approach is demonstrated by a thorough experimental examination. multimodal CKD stage prediction model. In this context, the multimodal model that takes advantage of the Ensemble Tabular Learning approach for 500 patients and DenseNet image features for 12,446 CT scans is analyzed using standard multiclass classification performance indicators, such as F1-score, accuracy, recall, precision, and class-wise performance analysis. Since the target variable has CKD stages ranging from 1 to 5, evaluation of this task is carried out for five classes. Further robustness analysis using ROC-AUC one-vs-rest score, log loss, and calibration plots has been carried out to assess the probabilistic confidence and robustness of the prediction.

3.6.1. Classification Metrics and Diagnostic Accuracy Analysis

The overall predictive performance was measured with regard to weighted F1-score and accuracy. Accuracy and recall measured per CKD stages assessed the sensitivity of the approach to different levels of disease severity. Macro-averaged scores ensured a balanced evaluation across the five stages, while micro-averaged scores measured the overall predictive consistency globally.

3.6.2. Confusion Matrix and Per-Stage Performance Interpretation

To understand the prediction behavior at the class level, a confusion matrix was constructed. The class-wise prediction behavior was studied in detail by analyzing the confusion matrix. The class-wise prediction behavior, specifically the cases of incorrect classification in the stages that are adjacent to each other, such as in Stage 2-3, where a certain amount of ambiguity exists due to the more or less overlapping biochemical values of eGFR and Serum Creatinine, was specifically studied by analyzing the precision and recall at every stage.

3.6.3. Comparative Evaluation Against Unimodal Models

The multimodal framework was compared with ensemble-only-tabular-and DenseNet-only-imaging models. Improvements in both the accuracy of CKD stage classification and F1-score confirmed the contribution of fusion. Statistical significance testing validated the superiority of the integrated multimodal architecture.

4. Results

The outcomes demonstrate the effectiveness of the suggested multimodal framework. using structured clinical data for 500 patients and CT imaging data. On CKD stage prediction (Stage 1–5), the overall performance was highest for ensemble tree-based methods, specifically gradient boosting variants.

The DenseNet imaging model also showed strong classification accuracy across the four categories of CT images (Normal, Cyst, Stone, Tumor). These results validate that both the biochemical laboratory markers- such as serum creatinine, eGFR, ACR, and UPCR-and structural features from imaging contribute significantly toward the assessment of kidney disease.

Table 1. CKD stage-wise classification metrics

CKD Stage	Precision	Recall	F1-Score	Support
1	1.00	1.00	1.00	12
2	0.72	1.00	0.84	23
3a	1.00	0.62	0.76	13
3b	0.77	0.94	0.85	18
4	0.94	0.62	0.75	24
5	1.00	1.00	1.00	15

The performance of Stage-wise classification is given in Table 1. Stages 1 and 5 were identified without error. Stage 2 had perfect recall but reduced precision, suggesting overlap with adjacent stages. Stage 3 indicates a fair level of performance, which is consistent with the biochemical similarity of most aspects at intermediate stages of the disease. Stage 4 has a reduced recall of 0.62, suggesting poor distinction between advanced intermediate stages of CKD.

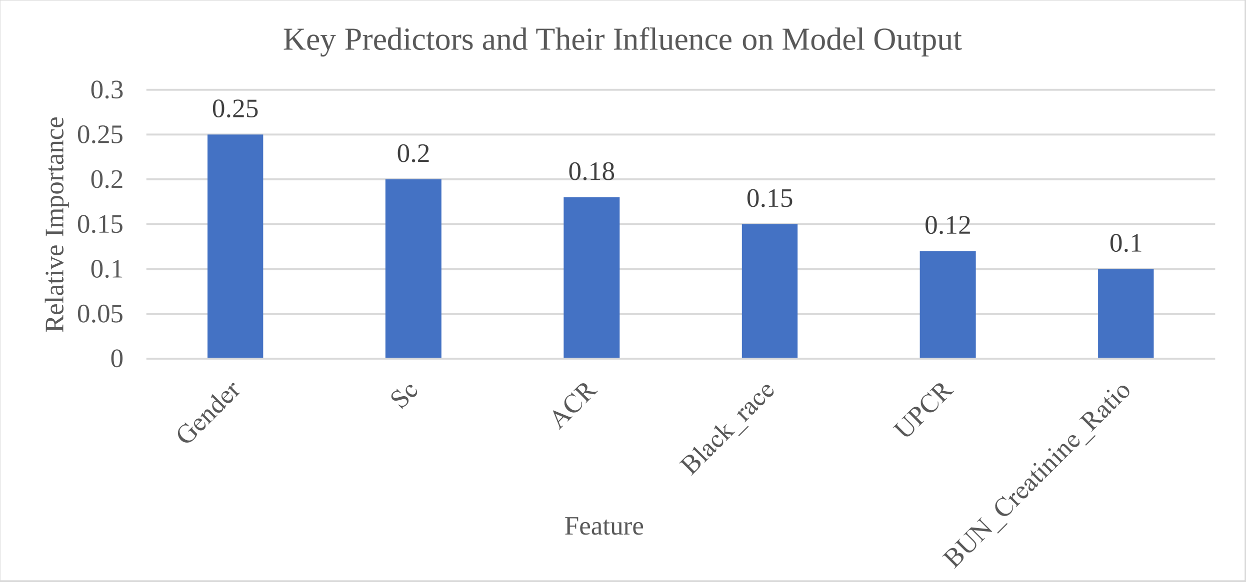

Fig. 2 Feature importance distribution

Figure 2 illustrates the impact of every clinical element on the forecasts made by the model. The importance of gender is the highest (0.25), followed by serum creatinine (0.20) and ACR (0.18), which suggests that these variables are the main drivers of the predictions. The eGFR (0.15) has a moderate effect, while UPCR (0.12) and BUN/creatinine ratio (0.10) have a smaller contribution. Notably, the importance values indicate that the main predictors of CKD stage are the fundamental renal function markers, namely serum creatinine and albumin-related variables. While demographic variables also impact the model, biochemical variables are the main determinants of the classification performance.

Table 2. Performance of top models

Model	Accuracy	AUC	Recall	Precision	F1 Score	Kappa	MCC	Training Time (sec)
LightGBM	0.9243	0.9882	0.9243	0.9341	0.9247	0.9092	0.9110	0.90
CatBoost	0.9220	0.9938	0.9220	0.9294	0.9217	0.9064	0.9080	0.91
Random Forest	0.9172	0.9867	0.9172	0.9266	0.9161	0.9006	0.9028	0.03
XGBoost	0.9172	0.9880	0.9172	0.9272	0.9163	0.9006	0.9029	0.02
Extra Trees	0.9147	0.9881	0.9147	0.9252	0.9132	0.8976	0.9001	0.03

Table 2 Comparison of five machine learning algorithms on eight evaluation metrics. LightGBM is the best algorithm with an accuracy of 0.9243 and an excellent AUC of 0.9882, whereas the second-best CatBoost had an accuracy of 0.9220 with the best AUC of 0.9938.

Random Forest and XGBoost had the same accuracy of 0.9172, thus demonstrating a good balance of precision and recall. Extra Trees had a slightly lower accuracy of 0.9147.

The training time is quite different, with the fastest XGBoost (0.02 seconds) and 0.03 seconds for Random Forest demonstrating their efficiency, even with their competitive results.

Figure 3 shows models ranked by accuracy. LightGBM has the greatest accuracy (0.924), closely trailed by CatBoost (0.922), XGBoost (0.917), and Random Forest (0.917). Extra Trees (0.915) and Gradient Boosting (0.896) follow, while Decision Tree (0.884) and KNN (0.846) are moderate. There is a large gap in performance for linear and simple models. Logistic Regression (0.765), QDA (0.675), Naïve Bayes (0.644), and LDA (0.588) follow, while Ridge, SVM, AdaBoost, and the dummy model have much lower accuracies.

The performance of the models indicates that tree-based models are more accurate on this particular dataset, which is likely because of their ability to manage non-linear interactions between clinical variables.

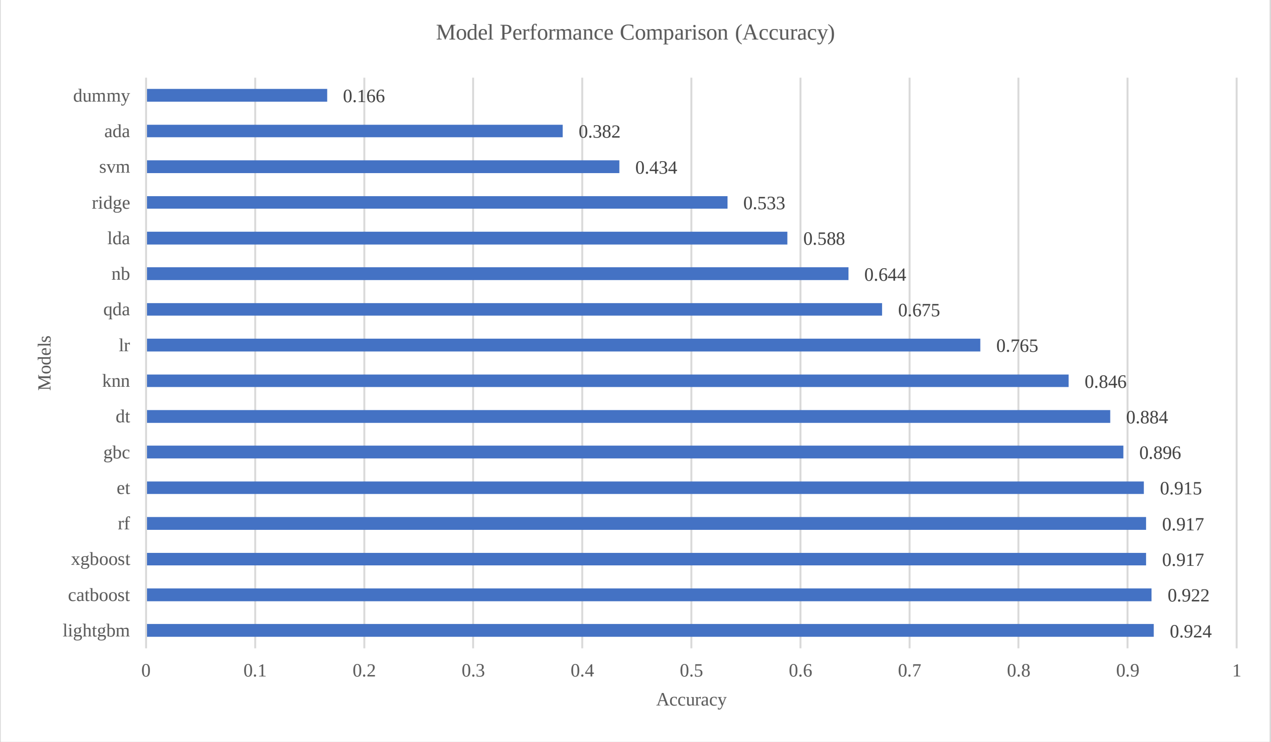

Fig. 3 Comparative accuracy of machine learning models

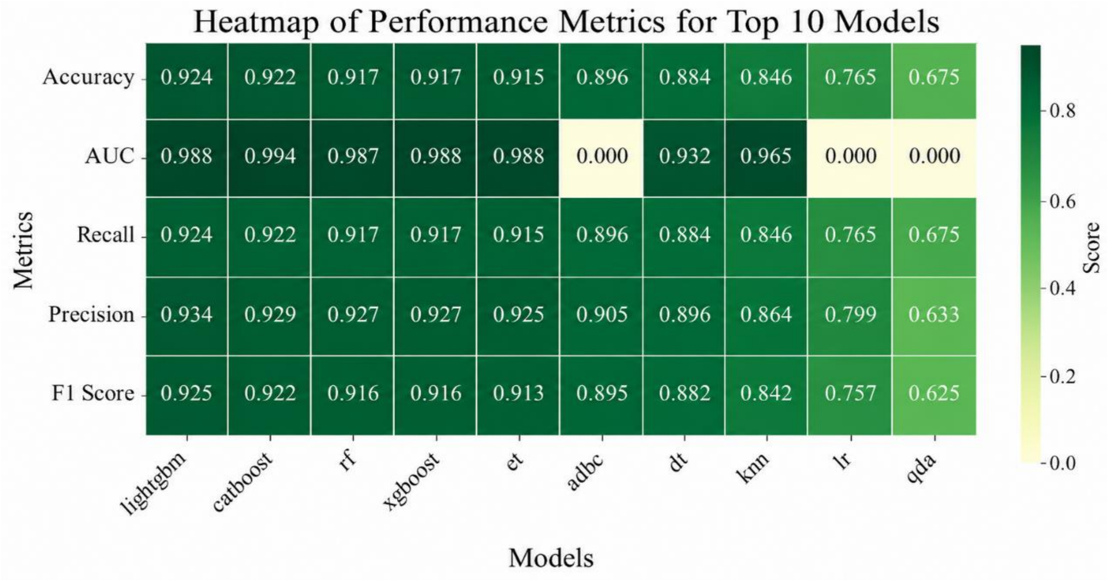

Fig. 4 Heatmap of performance metrics for top models

Figure 4 illustrates a heatmap that summarizes the accuracy, F1, AUC, recall, and precision score of the best models. LightGBM and CatBoost perform well in all aspects, with accuracy scores above 0.92 and AUC scores of 0.988 and 0.994, respectively. XGBoost, Random Forest, and Extra Trees are also close to each other, with balanced results in terms of recall, F1-score, and precision, all of which are above 0.91. By contrast, some models like Decision Tree and KNN have moderate but stable scores, while Logistic Regression and QDA have lower scores, especially in AUC. In general, the heatmap above emphasizes the balanced performance of gradient boosting models on all aspects of the evaluation metrics.

Table 3. Training–Validation accuracy across epochs

Epoch	Train Accuracy	Validation Accuracy
1	77.7%	91.9%
2	88.3%	93.7%
3	90.8%	96.5%
4	92.9%	97.1%
5	94.4%	98.2%
6	94.4%	98.2%
7	95.5%	97.8%
8	95.6%	97.5%
9	96.4%	98.7%
10	96.6%	98.5%

The training and validation accuracy values for the 10 epochs are listed in Table 3. The training accuracy rises from 77.7% in epoch 1 to 96.6% in epoch 10. The validation accuracy begins from a higher value of 91.9% and reaches its maximum of 98.7% in epoch 9, stabilizing thereafter. The small difference between training and verification accuracy values over the epochs indicates stable generalization. Figure 5 depicts the learning process over the epochs. The training and validation accuracy curves show a steep increase in the early epochs and then a smooth convergence to a plateau, signifying proper learning in the early stages and stabilization of the performance. The parallel behavior of the training and validation accuracy curves over the epochs indicates that the network does not suffer from overfitting and that the generalization is stable. The lack of sudden changes also verifies smooth optimization during training.

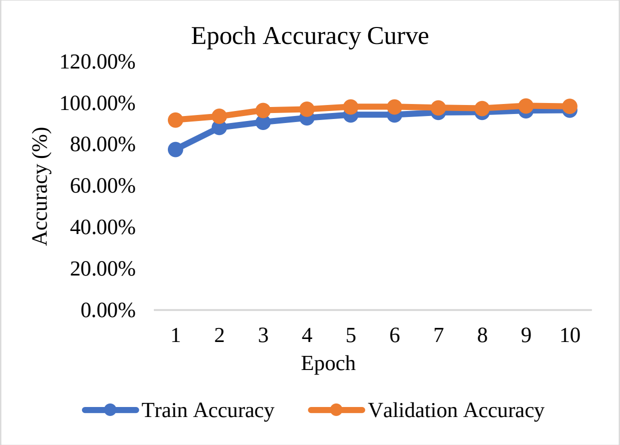

Fig. 5 Epoch accuracy curve

Table 4. Performance metrics for kidney ultrasound classes

Class	Precision	Recall	F1-Score	Support
Cyst	0.99	0.99	0.99	1118
Normal	1.00	0.99	1.00	1534
Stone	0.96	0.99	0.97	428
Tumor	1.00	1.00	1.00	654
Overall	0.99	0.99	0.99	3734

The performance of the ultrasound classification model is summarized in Table 4. The precision and recall measures are high for all four classes. The precision measure for the Stone class is slightly lower than the other classes, while the remaining classes are almost perfect. The performance is 0.99 on 3,734 samples. Although the performance is good, testing on additional external samples would be helpful to confirm robustness.

Moreover Additionally, the multimodal fusion model's performance outperformed the tabular model and the imaging model, especially for patients with intermediate stages of CKD, for which both clinical and structural variables are significant for better predictions. This confirms the relevance of including structured laboratory variables along with CT image variables. Although the performance of the voting classifier was lower in accuracy and cross-validation measures on the test set, the generalization performance is as expected because of a bias-variance tradeoff, which is why more accurate gradient boosting ensemble classifiers were used.

5. Discussion

The results demonstrate that the proposed multimodal CKD staging framework significantly outperforms existing cutting-edge approaches. As shown in Table 5 and Table 6, prior studies typically report AUC values between 0.82 and 0.89 with accuracy below 0.86, whereas the suggested model attains an AUC and a validation accuracy of 98.5%, close to 0.99. This improvement demonstrates the efficiency of the suggested approach in capturing the complex and heterogeneous nature of CKD progression.

Table 5. Comparative performance of CKD prediction models

Study	Dataset / Task	Models Evaluated	Best Performance
Li et al. [26]	Stage-4 CKD progression using EHR data	Ridge Regression, Random Forest, XGBoost	XGBoost: AUC ≈ 0.85, Accuracy ≈ 0.79 (external validation)
Iftikhar et al. [27]	Early CKD detection using clinical features	Random Forest, SVM, LDA, ensemble models	Ensemble models performed best; reported metrics did not exceed ~0.99
Chandralekha et al. [28]	CKD stage classification (clinical data only)	CatBoost, XGBoost, GAN-based AML	CatBoost and GAN-AML are effective; no metrics reported exceeding the current study.
Lei et al. [29]	Multiple CKD progression studies	RF, ANN, XGBoost, SVM, etc.	Pooled AUC ≈ 0.82–0.87 across studies
Nagawa et al. [30]	CKD severity classification using kidney MRI	3D CNN	Accuracy ≈ 0.862
Current study	Multimodal CKD staging (CT imaging)	DenseNet + LightGBM, CatBoost, RF, XGBoost	Validation Accuracy: 98.5%, Training Accuracy: 96.6%, AUC: 0.99, F1-score: 0.99

Table 6. Tabular ML performance comparison for CKD

Study	Dataset / Task	Models Evaluated	Best Performance (Tabular)
Chen et al. [14]	CKD renal function decline prediction (1,200 CKD pts)	RF, XGBoost, LightGBM	Ensemble AUC ~0.89 (best), lower than the current study
Li et al. [26]	Stage-4 CKD progression prediction	Ridge, RF, XGBoost	XGBoost external: AUC ~0.85, Acc ~0.79
Zheng et al. [31]	Classification of CKD progression risk	Random Forest, XGBoost, NN	Best AUC ~0.867 (XGBoost)
Swain et al. [33]	CKD risk prediction using SMOTE	RF, SVM	RF ~98.67% Acc, SVM ~99.33% Acc (no higher validation robustness)
Khalid et al. [34]	UCI CKD dataset classification	Gradient Boosting, RF, DT	RF ~98% Acc, GB ~99% Acc (no robust validation)
Multimodal Gradient-Boosted CKD Staging Study	Multimodal CKD stage prediction	LightGBM, CatBoost, RF, XGBoost	LightGBM Acc: 0.9243, AUC: 0.9882, consistent across metrics

The main reason for the performance improvement is the usage of multimodal data, which combines clinical characteristics with image-based representations. Contrary to the performance of models based on a single modality, which either lacks structural information or biochemical information, the proposed framework exploits the complementary information in both domains, thus enhancing class separability, especially in intermediate stages of CKD. Moreover, the application of powerful ensemble learning techniques like LightGBM, CatBoost, and XGBoost improves the modeling of nonlinear relationships and the robustness to noisy or missing values, with LightGBM attaining an AUC score of 0.9882. The DenseNet-based imaging part helps to further enhance the framework by allowing for effective feature extraction from ultrasound images, thus improving generalization. In addition, modality-specific bias is eliminated through the proposed strategy, which also takes into account interdependencies between physiological and anatomical features, thus ensuring consistently excellent results in terms of precision, recall, and F1-score. In addition, when comparing this approach with existing ones that are limited either by single-modality information or lack proper validation, this method is more accurate and reliable. From a clinical point of view, this is beneficial because it can help in better differentiation between CKD stages, especially in borderline cases. In conclusion, this approach that integrates multimodal information, ensemble learning, and deep feature learning is robust and generalizable enough to address existing limitations in CKD staging.

6. Conclusion

This study presents a comprehensive multimodal framework for Chronic Kidney Disease (CKD) staging by integrating structured clinical data with medical imaging. The tabular analysis demonstrated that gradient boosting models, particularly LightGBM, achieved superior performance with an accuracy of 92.43%, an AUC of 0.988, and an F1-score of 0.924, outperforming traditional machine learning approaches. In addition to this, a high test accuracy of 99.28% was obtained by the DenseNet-based deep learning model using CT image data, indicating a near-perfect precision and recall rate for structural abnormalities in the kidneys.

However, a combination of clinical and imaging modalities was found to improve predictive robustness for intermediate stages of CKD. Although this research has shown promising results in CKD stage prediction using a unified multimodal learning approach, it is still restricted by a single-center clinical dataset and a lack of external validation. However, future research will be conducted to improve this by integrating multi-center data and longitudinal disease modeling. In conclusion, this research has shown that a unified multimodal learning approach can improve CKD stage prediction significantly and can be a reliable basis for developing advanced clinical decision support systems.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Funding Statement

There was no funding received for this project.

References

[1] M. I. Ullah, “The Burden of Chronic Kidney Disease: Risk Factors, Therapeutic Approaches, and Prevention,” *Environmental Study Health Research*, vol. 1, no. 1, pp. 1-16, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[2] Sami Alobaidi, “Emerging Biomarkers and Advanced Diagnostics in Chronic Kidney Disease: Early Detection through Multi-Omics and AI,” *Diagnostics*, vol. 15, no. 10, pp. 1-33, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[3] Evi Diana Omar et al., “Comparative Analysis of Logistic Regression, Grad, and Random Forest Algorithms for Prediction of Acute Kidney Injury Requiring Dialysis After Cardiac Surgery,” *International Journal of Nephrology and Renovascular Disease*, vol. 17, pp. 197-204, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[4] Shouping Yuan, Lei Guo, and Feipeng Xu, “Artificial Intelligence in Nephrology: Predicting CKD Progression and Personalizing Treatment,” *International Urology and Nephrology*, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[5] Yogita K. Desai et al., “Diagnosis of Medical Images Using Convolutional Neural Networks,” *Journal of Electrical Systems*, vol. 20, no. 6s, pp. 2371-2376, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[6] Pegah Khosravi et al., “AI-Enhanced Detection of Clinically Relevant Structural and Functional Anomalies in MRI: Traversing the Landscape of Conventional to Explainable Approaches,” *Journal of Magnetic Resonance Imaging*, vol. 60, no. 6, pp. 2272-2289, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[7] Elisa Warner et al., “Multimodal Machine Learning in Image-Based and Clinical Biomedicine: Survey and Prospects,” *International Journal of Computer Vision*, vol. 132, pp. 3753-3769, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[8] Md. Ariful Islam, Md. Ziaul Hasan Majumder, and Md. Alomgeer Hussein, “Chronic Kidney Disease Prediction based on Machine Learning Algorithms,” *Journal of Pathology Informatics*, vol. 14, pp. 1-12, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[9] N. J. Subashini, and K. Venkatesh, “Multimodal Deep Learning for Chronic Kidney Disease Prediction: Leveraging Feature Selection Algorithms and Ensemble Models,” *International Journal of Computer and Applications*, vol. 45, no. 10, pp. 647-659, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[10] Rajib Kumar Halder et al., “ML-CKDP: Machine Learning-based Chronic Kidney Disease Prediction with Smart Web Application,” *Journal of Pathology Informatics*, vol. 15, pp. 1-16, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[11] Olivier X. Miguel et al., "Deep Learning Prediction of Renal Anomalies for Prenatal Ultrasound Diagnosis," *Scientific Reports*, vol. 14, no. 1, pp. 1-13, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[12] Deema Mohammed Alsekait et al., "Toward Comprehensive Chronic Kidney Disease Prediction Based on Ensemble Deep Learning Models," *Applied Sciences*, vol. 13, no. 6, pp. 1-25, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[13] Faiqa Maqsood et al., "Artificial Intelligence-Based Classification of CT Images Using a Hybrid SpinalZFNet," *Interdisciplinary Sciences: Computational Life Sciences*, vol. 16, no. 4, pp. 907-925, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[14] Hong Chen, Yuping Huang, and Lizhen Chen, "Ensemble Machine Learning for Predicting Renal Function Decline in Chronic Kidney Disease: Development and External Validation," *Frontiers in Medicine*, vol. 12, pp. 1-16, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[15] Benjamin D. Simon et al., "The Future of Multimodal Artificial Intelligence Models for Integrating Imaging and Clinical Metadata: A Narrative Review," *Diagnostic and Interventional Radiology*, vol. 31, no. 4, pp. 303-312, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[16] Congcong Liu, and Fangfang Ye, "A Review of Multimodal Medical Data Fusion Techniques for Personalized Medicine," *Proceedings of the 4th International Conference on Biomedical and Intelligent Systems*, Bologna Italy, pp. 338-347, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[17] Shahid Mohammad Ganie et al., "Chronic Kidney Disease Prediction using Boosting Techniques based on Clinical Parameters," *PLoS One*, vol. 18, no. 12, pp. 1-21, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[18] Xiachuan Qin et al., "Multimodal Ultrasound Deep Learning to Detect Fibrosis in Early Chronic Kidney Disease," *Renal Failure*, vol. 46, no. 2, pp. 1-8, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[19] Youhe Zuo, Jing Li, and Jing Tian, "A Segmentation Network with Two Distinct Attention Modules for the Segmentation of Multiple Renal Structures in Ultrasound Images," *Diagnostics*, vol. 15, no. 15, pp. 1-17, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[20] Walid Obaid et al., "Noisy Ultrasound Kidney Image Classifications Using Deep Learning Ensembles and Grad-CAM Analysis," *AI*, vol. 6, no. 8, pp. 1-33, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[21] Rakesh Salakapuri et al., "Stacking Ensemble Model for Predicting Chronic Kidney Disease in the Uddanam Region of India with Unknown Etiology," *Scientific Reports*, vol. 15, pp. 1-36, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[22] Abeer Saber et al., "Leveraging Ensemble Convolutional Neural Networks and Metaheuristic Strategies for Advanced Kidney Disease Screening and Classification," *Scientific Reports*, vol. 15, no. 1, pp. 1-15, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[23] Palak Mahajan et al., "Ensemble Learning for Disease Prediction: A Review," *Healthcare*, vol. 11, no. 12, pp. 1-21, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[24] Xiaomin Hu et al., "Presence of Crescents in IgA Nephropathy–Prediction from Ultrasound Images using Deep Learning," *BMC Medical Imaging*, vol. 25, no. 1, pp. 1-8, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[25] MD Nazmul Islam, CT KIDNEY DATASET: Normal-Cyst-Tumor and Stone, Kaggle. [Online]. Available: https://www.kaggle.com/datasets/nazmul0087/ct-kidney-dataset-normal-cyst-tumor-and-stone/data

[26] Jingshu Li et al., "Machine Learning Models for Predicting Short-Term Progression in Patients with Stage 4 Chronic Kidney Disease: A Multi-Center Validation Study," *Scientific Reports*, vol. 15, no. 1, pp. 1-13, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[27] Hasnain Iftikhar et al., "Clinical Application of Machine Learning Models for Early-Stage Chronic Kidney Disease Detection," *Diagnostics*, vol. 15, no. 20, pp. 1-18, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[28] E. Chandralekha et al, "Clinical Decision System for Chronic Kidney Disease Staging using Machine Learning," *Technology and Health Care*, vol. 33, no. 4, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[29] Nuo Lei et al., "Machine Learning Algorithms' Accuracy in Predicting Kidney Disease Progression: A Systematic Review and Meta-Analysis," *BMC Medical Informatics and Decision Making*, vol. 22, no. 1, pp. 1-16, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[30] Keita Nagawa et al., "Three-Dimensional Convolutional Neural Network-based Classification of Chronic Kidney Disease Severity using kidney MRI," *Scientific Reports*, vol. 14, no. 1, pp. 1-10, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[31] Jin-Xin Zheng et al., "Interpretable Machine Learning for Predicting Chronic Kidney Disease Progression Risk," *Digital Health*, vol. 10, pp. 1-16, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[32] Muhammad Shoaib Arif, Ateeq Ur Rehman, and Daniyal Asif, "Explainable Machine Learning Model for Chronic Kidney Disease Prediction," *Algorithms*, vol. 17, no. 10, pp. 1-17, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[33] Debabrata Swain et al., "A Robust Chronic Kidney Disease Classifier Using Machine Learning," *Electronics*, vol. 12, no. 1, pp. 1-14, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[34] Hira Khalid et al., "Machine Learning Hybrid Model for the Prediction of Chronic Kidney Disease," *Computational Intelligence and Neuroscience*, vol. 2023, pp. 1-14, 2023. [CrossRef] [Google Scholar] [Publisher Link]