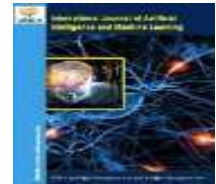




International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: <https://www.svedbergopen.com/>



Research Paper

Open Access

An Explainable Hybrid CNN–Machine Learning Framework for Accurate Chronic Kidney Disease Classification: A Comparative Study

D P Singh

Amity University Uttar Pradesh Greater Noida Campus. E-mail: drdps97@gmail.com, Orcid: <https://orcid.org/0000-0001-9494-4296>

Abstract

Chronic kidney disease (CKD) is a major global health burden, emphasizing the importance of early detection and timely management. Conventional machine-learning methods often depend on manually engineered features, whereas end-to-end deep-learning models may have limited interpretability when applied to structured clinical data. This study proposes a hybrid CNN–machine learning framework that combines one-dimensional Convolutional Neural Network (1-D CNN) feature learning with Random Forest (RF), Extreme Gradient Boosting (XGBoost), Support Vector Machine (SVM), and Logistic Regression (LR). The framework is evaluated using 400 CKD patient records containing 24 clinical variables, which are transformed into 34 model-ready features through missing-value treatment, categorical encoding, numerical standardization, and feature engineering. A stratified 80:20 train–test split and five-fold stratified cross-validation are used. The CNN generates a 64-dimensional feature representation for subsequent classification. Among the evaluated models, CNN–XGBoost achieves the best independent test performance, with 98.75% accuracy, 100% precision, 98% recall, 100% specificity, 98.99% F1-score, 0.9993 ROC-AUC, and 0.9996 PR-AUC. The confusion matrix shows 49 true positives, 30 true negatives, one false negative, and no false positives. SHAP analysis identifies serum creatinine, blood urea, hemoglobin, specific gravity, albumin, and packed cell volume as important predictors. Overall, combining CNN-based feature learning with XGBoost provides a potentially effective and interpretable approach to CKD classification and data-driven clinical decision-making. External validation using larger, multicentre, prospective datasets remains necessary to establish its clinical applicability.

Keywords: Chronic Kidney Disease; CKD Classification; Deep Learning; 1-D Convolutional Neural Network; Machine Learning; CNN–XGBoost; XGBoost; Clinical Prediction; Explainable Artificial Intelligence; SHAP; Feature Learning; Clinical Decision Support; Medical Artificial Intelligence.

1. Introduction

Chronic kidney disease (CKD) is a progressive condition characterized by abnormalities in kidney structure or function that persist for at least three months and is classified according to its cause, glomerular filtration rate (GFR), and albuminuria (KDIGO)[30]. With an estimated global burden of about 850 million people, CKD represents a major public health concern and is associated with cardiovascular complications, kidney failure, hospitalization, impaired quality of life, and premature death (ISN)[25]. Because the disease can progress silently, particularly during its early stages, reliable methods for early identification and risk assessment are needed. Diabetes and hypertension are important contributors to CKD, and progressive loss of renal function may eventually necessitate dialysis or transplantation (KDIGO)[30]. Delays in diagnosis can reduce opportunities for preventive management; consequently, artificial intelligence (AI) methods that use laboratory and imaging information may facilitate earlier detection and risk stratification [43]. Conventional assessment draws on demographic, clinical, biochemical, and laboratory measurements, whose nonlinear and interdependent relationships may be difficult to model adequately using traditional techniques. AI and ML are therefore receiving increasing attention for identifying complex patterns relevant to CKD diagnosis, classification, and progression prediction [23,43].

The use of artificial intelligence (AI) and machine learning (ML) has broadened the possibilities for CKD detection, diagnosis, risk estimation, and clinical decision support by enabling the analysis of patterns within electronic health records, laboratory results, physiological measurements, medical images, and related clinical information [23]. Methods such as Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), and XGBoost have been applied to CKD classification and progression-risk prediction, while explainability techniques such as SHAP and LIME can help clarify model outputs [37]. XGBoost has also demonstrated strong discriminatory performance, indicating the usefulness of nonlinear ensemble methods for heterogeneous clinical datasets[39]. Since LR, SVM, RF, and XGBoost represent different modelling principles, evaluating them together may reveal benefits that are not apparent from using a single classifier.

Deep-learning approaches offer an alternative by automatically learning feature representations from input data. While convolutional neural networks are predominantly used for image-based tasks, 1-D CNNs can also process structured or sequential clinical variables to capture local dependencies and higher-order patterns, generating compact representations for subsequent classification. AI-based approaches incorporating laboratory and imaging data have demonstrated potential for early CKD detection and risk assessment [43]. Therefore, CNN-based feature learning may identify complex feature interactions that are difficult to define manually, while conventional ML classifiers can subsequently evaluate the predictive value of the learned representations.

Despite continuing progress in CKD prediction, important shortcomings remain in existing ML and deep-learning studies. These include reliance on small or single-centre datasets, uncertain generalizability, dependence on preprocessing and model settings, handcrafted feature construction, overfitting, limited interpretability, insufficient stage-specific prediction, privacy issues, and reproducibility concerns [23,37,39]. Classical ML methods are well suited to structured clinical data, while deep networks can learn nonlinear representations automatically; however, the two approaches are frequently studied as separate alternatives. A key unresolved issue is therefore whether CNN-derived representations can be systematically combined with diverse conventional classifiers. Guided by the premise that learned nonlinear representations may encode clinically relevant CKD patterns more effectively, the proposed framework uses a 1-D CNN followed by RF, XGBoost, SVM, and LR to examine whether the learned representation consistently improves classification across different ML paradigms.

To address this gap, the study introduces a hybrid CNN-ML framework for CKD classification and prediction, in which a 1-D CNN learns representations from pre-processed clinical variables and four conventional classifiers—RF, XGBoost, SVM, and LR—perform the final classification. The framework evaluates the models using accuracy, precision, recall/sensitivity, specificity, F1-score, ROC-AUC, cross-validation results, and robustness analyses, while explainability analysis is used to identify clinically important predictors and improve transparency. Because each classifier is evaluated using the same CNN-generated feature space, the study seeks to determine which downstream model is most effective and robust, whether nonlinear learned representations offer an advantage over conventional approaches, and how the framework can support future research on explainability, external validation, stage-wise prediction, and clinical decision support.

2. Related Work

CKD requires timely identification, and the increasing availability of electronic health records, laboratory measurements, demographic information, and imaging has encouraged the use of ML and deep learning (DL) for detection and risk prediction. Prior investigations have employed conventional ML, DL, and hybrid strategies; however, a systematic review of 68 AI/ML studies reported persistent concerns regarding dataset comparability, generalizability, external validation, interpretability, and clinical assessment [27]. The present work builds on this literature by combining CNN-based automatic feature learning with RF, XGBoost, SVM, and LR, thereby bringing learned nonlinear representations together with complementary conventional classification approaches.

For structured CKD datasets, traditional algorithms including LR, SVM, RF, Decision Tree, KNN, and gradient-boosting methods—have been extensively investigated because they can accommodate heterogeneous clinical variables without excessive computational demands. Debal and Sitote [21] reported 99.8% accuracy for binary CKD classification using RF after recursive feature elimination and also emphasized the importance of extending prediction to multiple CKD stages. Comparative work has shown that model performance can vary according to the dataset and prediction objective [16]. RF can capture nonlinear relationships and provide feature-importance measures, SVM is useful for complex or high-dimensional spaces, and LR offers an interpretable linear reference model. Ensemble strategies can also benefit from combining different classifiers; for example, Chhabra et al. [28] reported 99.5% cross-validated accuracy with an SVM–RF–AdaBoost ensemble using SMOTE and k-fold validation. Nevertheless, conventional ML results can be strongly influenced by preprocessing decisions, scaling, feature selection, missing-data treatment, feature engineering, and class imbalance, and these models generally operate on features prepared in advance.

Deep learning offers a different strategy by learning useful representations from the input rather than depending entirely on manually constructed features. Although CNNs originated primarily in image analysis, they can also be adapted to structured or sequential clinical variables to model nonlinear dependencies with less manual feature construction. Ma et al. [18] investigated deep-learning-based CKD detection using medical imaging within an IoMT-oriented framework, while Saif et al. [32] introduced Deep-kidney, integrating CNN, LSTM, and deep-ensemble components for CKD prediction and reporting accuracies above 99% in their experimental setting. Nagawa et al. [31] also showed that 3-D CNNs can classify CKD severity from bilateral kidney MRI data. Such models, however, require adequate sample sizes, appropriate architecture selection, regularization, and careful validation. Small datasets and imbalanced

classes may increase overfitting and weaken the recognition of minority cases, while the relative lack of interpretability can hinder clinical adoption. Sanmarchi et al. [27] further noted that generalizability across populations was infrequently assessed and highlighted interpretability, fairness, generalizability, and clinical validation as continuing challenges.

Hybrid CNN-ML strategies seek to exploit the complementary strengths of CNN feature learning and conventional classifiers such as SVM, RF, XGBoost, and LR. In CKD applications, the CNN can transform heterogeneous clinical variables into nonlinear representations, while the downstream classifier determines the final class using its own decision mechanism. Ramu et al. [40] demonstrated the feasibility of a CNN-SVM arrangement for CKD, reporting 96.8% accuracy and an improvement over standalone SVM and RF, although additional computational cost was noted. Evaluating several classifiers on an identical CNN-derived representation can therefore help establish whether any observed benefit is general across classifiers or primarily dependent on the choice of downstream model.

The current CKD-AI literature continues to show several limitations, including the use of small or benchmark datasets and restricted generalizability [27], a predominance of binary rather than stage-specific prediction [21], dependence on manual preprocessing and feature construction, susceptibility to overfitting, and interpretability difficulties in deep-learning systems. Ramu et al. [40] examined CNN-SVM, whereas Saif et al. [32] investigated deep-ensemble approaches; however, systematic testing of several heterogeneous classifiers using one common CNN representation remains comparatively limited. Direct comparison among published studies is additionally complicated by differences in datasets, preprocessing procedures, data splitting, validation strategies, and evaluation measures [27].

This study therefore focuses on the controlled integration of CNN feature learning with RF, XGBoost, SVM, and LR within a shared learned feature space. The design brings together tree-based ensemble learning, boosting, margin-based classification, and linear probabilistic modelling to determine whether a common CNN representation consistently improves CKD classification and prediction. The framework has four main objectives: to determine whether CNN-based representation learning improves CKD classification; to compare RF, XGBoost, SVM, and LR when all receive the same learned features; to examine the relative value of nonlinear and ensemble classifiers compared with a linear model; and to evaluate the hybrid systems using a broad set of performance measures. In this way, the study connects deep representation learning with conventional clinical ML and seeks to identify a model that is both effective and sufficiently robust for further CKD prediction research.

3. Materials and Methods

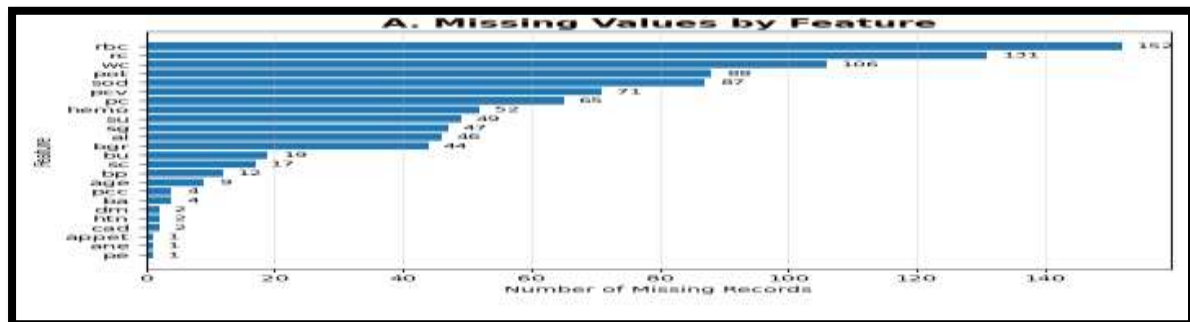
A hybrid CNN-ML methodology is developed for CKD classification and prediction by combining 1-D CNN-based feature extraction with RF, XGBoost, SVM, and LR classifiers. Model performance is evaluated using multiple performance measures and stratified cross-validation, while SHAP is employed to identify the clinical variables that most strongly influence the predictions, thereby supporting the goal of developing interpretable and potentially generalizable CKD models [23,37,39,43].

3.1 Dataset and Pre-processing: In this study, we used the Chronic Kidney Disease dataset from the UCI Machine Learning Repository to develop and evaluate the proposed CKD classification framework (UCI Machine Learning Repository) [12]. The analysis is conducted on 400 CKD records containing 24 clinical predictors. Median/mode imputation, categorical encoding, and normalization are applied during preprocessing, producing 34 features suitable for the 1-D CNN and subsequent ML classifiers (International Society of Nephrology [ISN], KDIGO) [25,30].

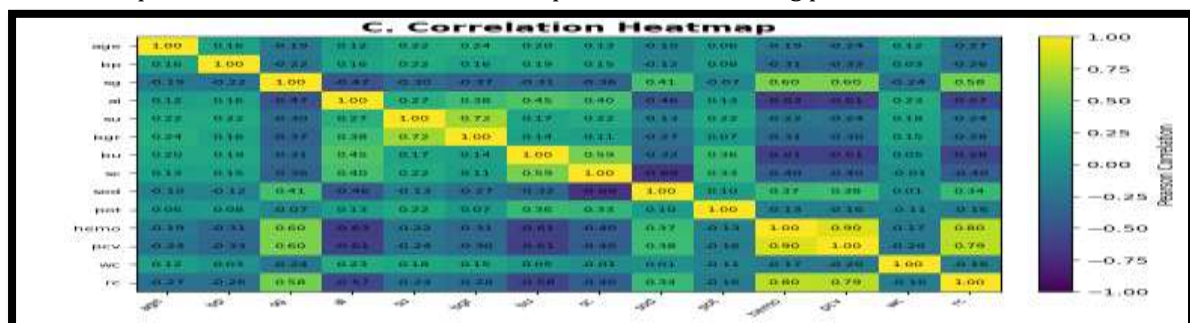
3.1.1 Identification and Clinical Features: The dataset included 14 quantitative and 10 qualitative predictors. Missing or invalid values were handled using median imputation for numerical variables and mode imputation for categorical variables, followed by appropriate encoding[20]. Non-numeric and inconsistent entries were standardized before analysis, and categorical variables were converted into numerical form. The numerical variables were then standardized using the z-score transformation: $z_i = \frac{x_i - \mu}{\sigma}$, where x_i denotes the observed feature value, μ represents the mean calculated from the training data, and σ denotes the corresponding standard deviation[26].



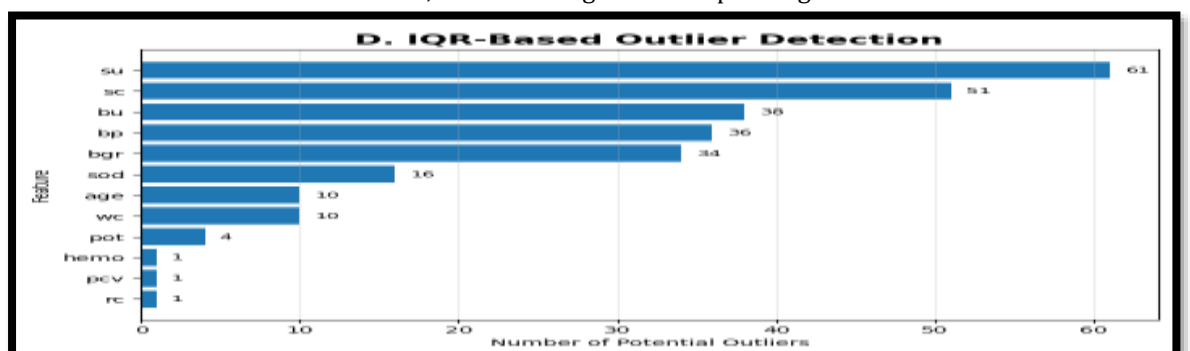
Numerical missing observations are replaced with the corresponding training-set median, whereas missing categorical entries were filled using the training-set mode.



3.1.2 Correlation Heatmap: The correlation analysis shows a very strong positive association between hemoglobin and packed cell volume ($r = 0.90$), along with a strong positive association between hemoglobin and red blood cell count ($r = 0.80$). A moderate positive association is observed between specific gravity and serum glucose ($r \approx 0.60$). In contrast, sodium shows a weak negative association with blood urea ($r = -0.32$), while packed cell volume shows a moderate negative association with blood urea ($r = -0.61$). Overall, these correlations indicate that several clinical variables are interrelated in the CKD dataset and provide useful context for the subsequent feature-learning process.

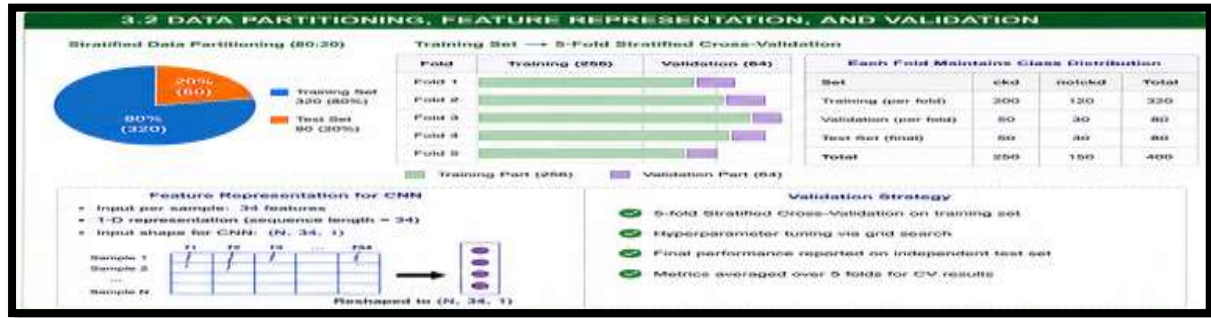


To prevent information leakage, all preprocessing parameters are estimated exclusively from the training data within each cross-validation fold, without using the corresponding validation data.

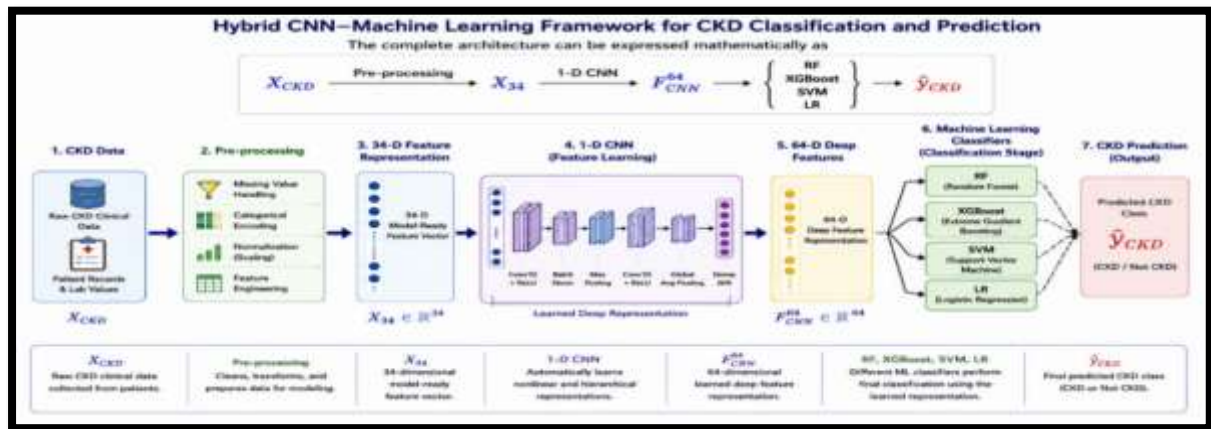


3.2 Data Partitioning, Feature Representation, and Validation: The 400 observations are stratified into 320 training records and an independent test set of 80 records, while maintaining the CKD/non-CKD class distribution. Five-fold stratified cross-validation is performed on the training data, whereas the test set is kept completely separate throughout model development and parameter selection [11]. Following preprocessing and feature transformation, each observation is represented by 34 model-ready variables,

which are arranged as a single-channel input for the 1-D CNN. The network then learns a compact nonlinear representation, reducing dependence on manually specified feature interactions [9].



3.3 Proposed CNN-Machine Learning Framework: The proposed architecture integrates 1-D CNN-based feature learning with four downstream machine-learning classifiers: Random Forest (RF), XGBoost (XGB), Support Vector Machine (SVM), and Logistic Regression (LR). In this framework, feature learning and classification are performed as two separate stages. First, the CNN extracts a compact feature representation from the pre-processed clinical variables. The learned features are then independently provided to each ML classifier for final CKD prediction. The overall framework can be mathematically expressed as follows:



3.3.1 1-D CNN Feature-Learning Module: The ReLU activation function introduces nonlinearity into the CNN, enabling learning of complex clinical patterns and supporting effective gradient-based learning [4]. It transforms the learned feature representations before they are passed to the downstream machine-learning classifiers. Each transformed clinical record is provided to the CNN as an input tensor with the shape: $X \in \mathbb{R}^{34 \times 1}$. For a convolutional layer l , the output at position i can be expressed as: $h_i^l = \sigma(\sum_{k=0}^{K-1} w_k^{(l)} x_{i+k}^{(l-1)} + b^l)$, where $x^{(l-1)}$ is the input representation from the preceding layer, $w_k^{(l)}$ denotes the convolutional kernel weights, K is the kernel size, b^l is the bias, and $\sigma(\cdot)$ represents the activation function [9].

The proposed 1-D CNN architecture consists of 64 filters in the initial Conv1D layer and 128 filters in the subsequent layer, with both convolutional layers using a kernel size of 3. Batch normalization is incorporated to maintain stable intermediate feature representations [10], whereas the ReLU activation function enables the network to model nonlinear relationships:

$$ReLU(z) = \max(0, z)$$

Max-pooling compresses the intermediate feature maps by reducing their spatial dimensions while preserving the most informative learned features. Subsequently, the convolutional and pooling layers are followed by Global Average Pooling (GAP), which summarizes the information contained within each feature map: $g_j = \frac{1}{N} \sum_{i=1}^N h_{ij}$, where h_{ij} represents the activation at position i of feature map j , and N is the number of spatial positions. The resulting representation is passed through a fully connected layer containing 64 neurons:

$$F_{CNN} = ReLU(W_d g + b_d), \text{ with } F_{CNN} = \mathbb{R}^{64}.$$

Accordingly, the CNN transforms the initial 34-feature clinical input into a concise 64-dimensional learned representation, which is subsequently used as input for the downstream machine-learning classifiers [7,9].

3.3.2 CNN-Random Forest: The first hybrid branch uses the 64-dimensional CNN representation as input to a Random Forest classifier: $F_{CNN}^{64} \xrightarrow{RF} \hat{Y}_{RF}$.

Random Forest constructs an ensemble of decision trees using bootstrap sampling and randomized feature selection. The final prediction is determined from the aggregated predictions of the individual trees (Breiman, 2001): $\hat{Y}_{RF} =$

$\text{mod}\{T_1(\mathbf{F}), T_2(\mathbf{F}), \dots, T_M(\mathbf{F})\}$, where T_m denotes the m -th decision tree and M is the number of trees. This architecture enables the CNN to learn the feature representation while Random Forest models nonlinear relationships and feature interactions in the resulting representation [2].

3.3.3 CNN-XGBoost: The second hybrid branch employs Extreme Gradient Boosting (XGBoost) on the learned CNN representation: $F_{CNN}^{64} \xrightarrow{XGBoost} \hat{y}_{XGB}$.

XGBoost constructs an additive ensemble of decision trees, with successive trees designed to minimize the remaining prediction error [13]. The general prediction function can be represented as $\hat{y}_i = \sum_{m=1}^M f_m(\mathbf{F}_i)$, $f_m \in F$, where f_m represents the m -th regression tree and F denotes the space of decision trees.

The corresponding regularized objective can be expressed as $\mathcal{L} = \sum_i l(y_i, \hat{y}_i) + \sum_{m=1}^M \Omega(f_m)$, where $l(y_i, \hat{y}_i)$ is the prediction loss and $\Omega(f_m)$ controls model complexity [13]. Thus, the CNN performs automated representation learning, whereas XGBoost performs nonlinear ensemble classification over the learned feature space.

3.3.4 CNN-SVM: The third branch uses a Support Vector Machine (SVM) with the 64-dimensional CNN representation: $F_{CNN}^{64} \xrightarrow{SVM} \hat{y}_{SVM}$.

SVM determines a decision boundary that maximizes the separation between the two classes [1]. For a nonlinear kernel formulation, the decision function can be written as $f(\mathbf{F}) = \sum_{i=1}^n \alpha_i y_i K(\mathbf{F}_i, \mathbf{F}) + b$, where α_i represents the learned support-vector coefficients, y_i is the class label, $K(\cdot, \cdot)$ is the kernel function, and b is the bias term.

An RBF kernel is employed: $K(\mathbf{F}_i, \mathbf{F}) = \exp(-\gamma \|\mathbf{F}_i - \mathbf{F}\|^2)$, where γ controls the influence of individual training observations. This allows nonlinear decision boundaries to be constructed within the learned CNN feature space [1].

3.3.5 CNN-Logistic Regression: The fourth hybrid branch uses Logistic Regression as a comparatively simple and interpretable classifier over the learned CNN representation:

$$F_{CNN}^{64} \xrightarrow{LR} \hat{y}_{LR}$$

For each observation, the probability of belonging to the positive CKD class is obtained through the logistic transformation:

$$P(y = 1|\mathbf{F}) = \frac{1}{1 + \exp[-(\beta_0 + \beta^T \mathbf{F})]}, \text{ where } \mathbf{F} \in \mathbb{R}^{64} \text{ is the CNN-derived representation, } \beta \text{ is the vector of learned coefficients, and } \beta_0 \text{ is the intercept [6].}$$

The predicted class can subsequently be defined as $\hat{y}_{LR} = \begin{cases} 1, & P(y = 1|\mathbf{F}) \geq \tau, \\ 0, & P(y = 1|\mathbf{F}) < \tau, \end{cases}$ where τ denotes the predefined classification threshold. Logistic Regression therefore provides a useful downstream linear classifier for determining whether the CNN representation itself contains sufficient discriminatory information [6].

3.3.6 Hybrid Prediction and Comparative Evaluation: Each downstream classifier independently produces a CKD probability: $P_{RF}, P_{XGB}, P_{SVM}, P_{LR}$.

The corresponding class predictions are obtained using the same predefined classification threshold: $\hat{y}_i = I(P_m \geq \tau)$, $m \in \{RF, XGB, SVM, LR\}$, where $I(\cdot)$ is the indicator function.

The four hybrid configurations were assessed using a consistent evaluation protocol that included accuracy, precision, recall (sensitivity), specificity, F1-score, ROC-AUC, PR-AUC, and confusion-matrix analysis [3,11]. Applying the same evaluation criteria across all models allowed a controlled comparison of the predictive contribution of the 64-dimensional feature representation learned by the CNN across the different machine-learning classifiers.

The overall mathematical formulation of the proposed framework can therefore be summarized as $\hat{y}_m = C_m[CNN(X_{34})]$, $m \in \{RF, XGB, SVM, LR\}$, where $CNN(X_{34}) = F_{CNN}^{64}$ represents the learned deep feature space and C_m denotes one of the four downstream classifiers. This formulation explicitly separates feature learning from classification, which constitutes the central methodological principle of the proposed hybrid CNN-ML framework [3,11].

3.3.7 Evaluation metrics: The distribution of correct and incorrect predictions in the independent test set was used to characterize classification quality through measures such as accuracy, precision, recall, and class-specific performance [33–36,41,42].

$$3.3.7.1 \text{ Accuracy: } Accuracy = \frac{TP + TN}{TP + TN + FP + FN},$$

$$3.3.7.2 \text{ Precision: } Precision = \frac{TP}{TP + FP}$$

$$3.3.7.3 \text{ Recall: } Recall = \frac{TP}{TP + FN}$$

$$3.3.7.4 \text{ Specificity: } Specificity = \frac{TN}{TN + FP}$$

Where TP= True positives, TN= True negatives, FP= False positives and FN= False negatives.

3.3.8 Five-Fold Stratified Cross-Validation: Five-fold stratified cross-validation was employed to assess model performance while maintaining approximately similar class proportions across the individual folds. This approach is particularly appropriate for classification tasks where class imbalance may influence the evaluation process [5]. For five-fold stratified cross-validation, each performance measure was computed separately for every fold and then summarized using the corresponding mean and standard deviation.

The final 80-observation test set was not used during model selection or tuning and was evaluated only after the selected model had been trained, providing an independent estimate of out-of-sample performance. Where the mean performance across the five folds is $\bar{M} = \frac{1}{5} \sum_{i=1}^5 M_i$ and the inter-fold standard

deviation is $SD = \sqrt{\frac{1}{4} \sum_{i=1}^5 (M_i - \bar{M})^2}$

The performance variability across folds was summarized using the mean and standard deviation. SHAP (SHapley Additive explanations) was subsequently used to quantify the contribution of individual clinical variables to model predictions, providing a consistent approach to interpreting the resulting ML model [14].

3.3.9 Model Training: CNN parameters were learned with the Adam optimizer by minimizing binary cross-entropy (BCE) loss. For each training observation, the network produces an estimated probability of CKD, and the loss measures the discrepancy between these probabilities and the observed binary labels. Where: N = Number of training observations, y_i = True CKD class for observation i , p_i = Predicted probability of CKD for observation i , θ = trainable CNN parameters.

Adam updates the trainable parameters iteratively using the learning rate η and bias-corrected first- and second-moment estimates of the gradients [8], together with the small constant ϵ used for numerical stability. The training objective is therefore to reduce the difference between observed CKD outcomes and the probabilities predicted by the CNN.

$$\theta^* = \arg \min_{\theta} \left\{ -\frac{1}{N} \sum_{i=1}^N [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)] \right\}$$

In practical terms, this objective encourages the network to produce probabilities that closely correspond to the observed CKD labels [8].

4. Experimental Results

The experimental analysis compares conventional ML baselines, a standalone 1-D CNN, and four hybrid CNN-ML configurations on the CKD dataset. Accuracy, precision, recall, specificity, F1-score, ROC-AUC, PR-AUC, and confusion matrices were considered, while five-fold cross-validation and alternative imputation analyses were used to examine consistency and robustness.

4.1 Dataset Characteristics and Baseline Performance: The dataset contained 400 observations, 24 clinical predictors (14 numerical and 10 categorical), and two outcome groups: CKD (62.5%) and non-CKD (37.5%).

A. Dataset Characteristics		B. Missing Value Summary		
Source	CKD Clinical Dataset (UCI)	Total missing entries in predictor features: 1,012		
File	Chronic_Key_Disease_dataset.csv	Missing Count	% Missing	Interpretation
Total Samples	400	0	12 (50.0%)	Complete
Total Features	26 (24 predictors + 1 target + 1 id)	1 – 10	6 (25.0%)	Low Missing
Predictor Features	24	11 – 50	4 (16.7%)	Moderate Missing
Numerical Features	14	> 50	2 (8.3%)	High Missing
Categorical Features	10			
Target Variable	classification			
Classes	CKD, Not CKD			

The class distribution corresponds to an approximately 1.67:1 imbalance. Across the predictor variables, 1,012 missing entries were recorded; the largest counts occurred for pc (364), pcc (167), and ba (143). These missing values were addressed during the preprocessing stage.



Among the conventional ML baselines, XGBoost produced the highest accuracy (97.80%) and ROC-AUC (0.9990), whereas RF, SVM, and LR each recorded 97.50% accuracy. These baseline results establish the reference point against which the CNN-derived and hybrid representations were assessed.

Model	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F1-Score (%)	ROC-AUC	PR-AUC
Random Forest	97.50	98.00	98.00	96.67	98.00	0.9986	0.9987
XGBoost	97.50	98.00	98.00	97.33	98.00	0.9980	0.9981
SVM (RBF)	97.50	98.00	98.00	96.67	97.96	0.9981	0.9984
Logistic Regression	97.50	98.00	98.00	98.67	97.96	0.9979	0.9982

4.2 CNN and Hybrid CNN-ML Performance: The 1-D CNN learned nonlinear representations from the 34 processed variables, and these representations were subsequently classified by RF, SVM, LR, and XGBoost. The standalone CNN produced 97.50% accuracy, a 97.96% F1-score, ROC-AUC of 0.9981, and PR-AUC of 0.9984. The strongest hybrid configuration was CNN-XGBoost, with 98.75% accuracy, 100% precision, 98% recall, 100% specificity, 98.99% F1-score, 0.9993 ROC-AUC, and 0.9996 PR-AUC. CNN-RF, CNN-SVM, and CNN-LR each reached 97.50% accuracy, with ROC-AUC values of 0.9986, 0.9981, and 0.9979, respectively.

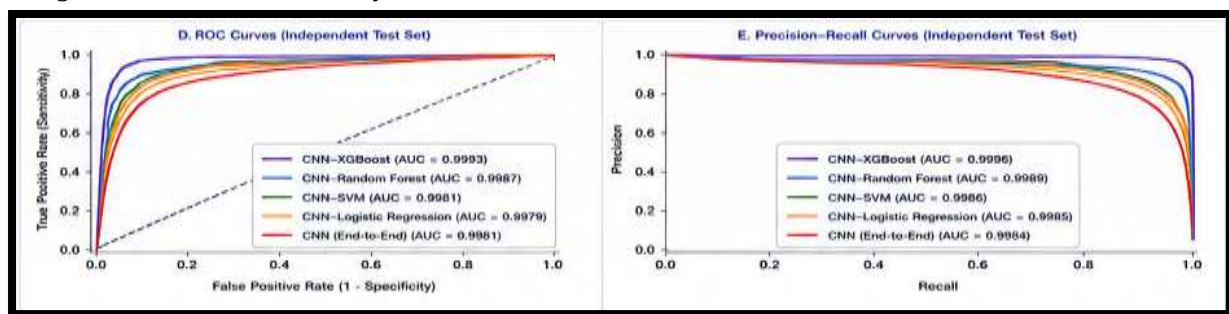
Metric	Value (%)
Accuracy	97.50
Precision	98.00
Recall (Sensitivity)	98.00
Specificity	96.00
F1-Score	97.96
ROC-AUC	0.9981
PR-AUC	0.9984

Model	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F1-Score (%)	ROC-AUC	PR-AUC
CNN-Random Forest	97.50	98.00	98.00	96.67	98.00	0.9986	0.9988
CNN-SVM (RBF)	97.50	100.00	96.00	100.00	97.96	0.9981	0.9986
CNN-Logistic Regression	97.50	100.00	96.00	100.00	97.96	0.9979	0.9984
CNN-XGBoost (Proposed)	98.75	100.00	98.00	100.00	98.99	0.9993	0.9996

Best Hybrid Model: CNN-XGBoost
Achieved the highest performance across all metrics.

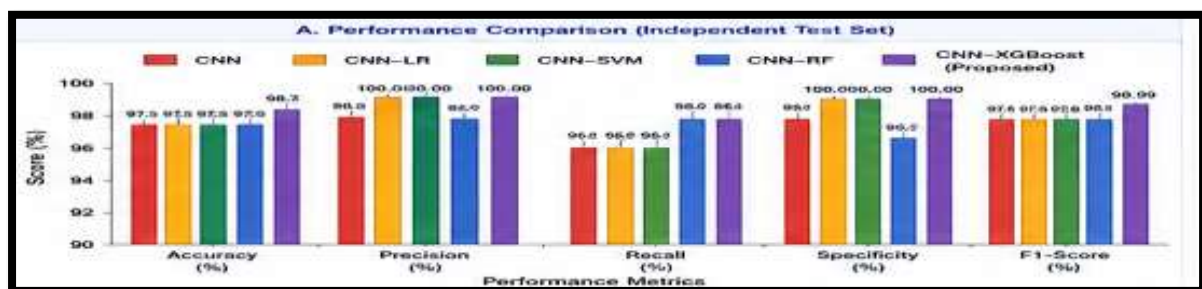
The proposed CNN-XGBoost model generalizes best on unseen data.

Both ROC and precision-recall curves indicated strong discriminatory ability for the evaluated models. CNN-XGBoost produced the largest ROC-AUC and PR-AUC values, making it the best-performing configuration within the tested hybrid framework.

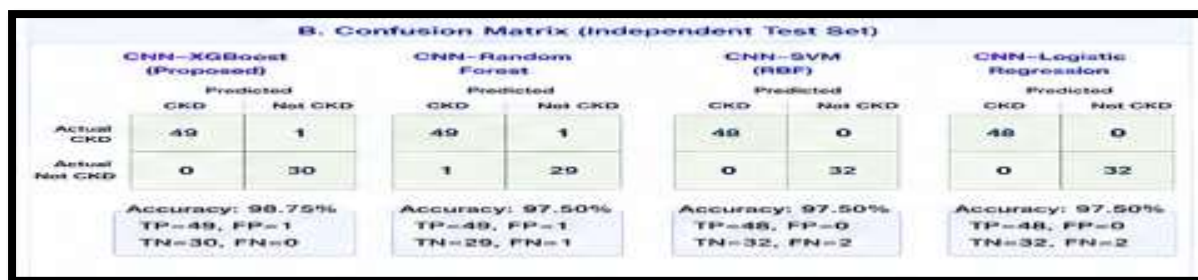


4.3 Comparative, Classification, and Robustness Analysis

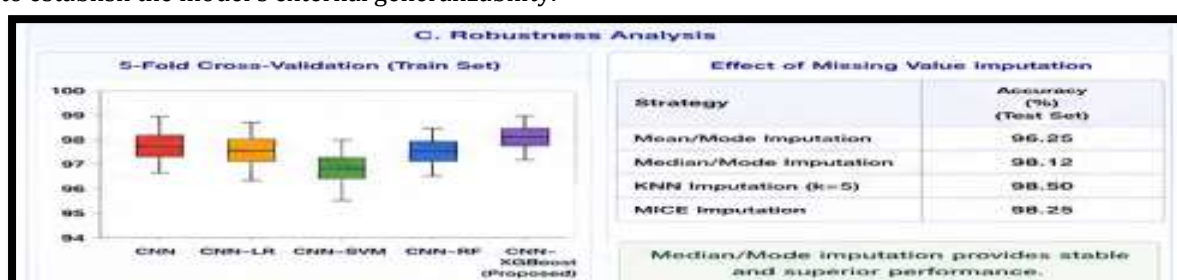
4.3.1 Comparative performance analysis: Performance remained strong across the evaluated measures, with CNN-XGBoost occupying the top position. Its 100% specificity and precision together with 98% recall indicate effective separation of CKD and non-CKD observations. The resulting order is CNN-XGBoost, CNN-RF, CNN-SVM, and CNN-LR, suggesting that the value of the learned CNN representation depends to some extent on the downstream classifier used with it.



4.3.2 Confusion Matrix and Classification Performance: For CNN-XGBoost, the confusion matrix for the independent test set contained 49 TP, 1 FN, 0 FP, and 30 TN, corresponding to 79 correctly classified observations out of 80 and an accuracy of 98.75%. CNN-RF achieved an accuracy of 97.50%, with 49 TP, 1 FN, 1 FP, and 29 TN. CNN-SVM and CNN-LR each achieved 97.50% accuracy, with 48 TP, 2 FN, 0 FP, and 30 TN. The absence of false-positive predictions and the low number of false-negative predictions for CNN-XGBoost indicate strong class separation on the independent test set.



4.3.3 Robustness and Sensitivity Analysis: Robustness is assessed using five-fold stratified cross-validation, alternative missing-data imputation strategies, independent testing, confusion-matrix analysis, ROC-AUC, and PR-AUC. Test accuracy remains high across the different imputation approaches, with accuracies of 96.25% for mean/mode, 98.12% for median/mode, 98.50% for KNN, and 98.25% for MICE. Median/mode imputation is retained as the reference configuration. Although these findings indicate strong performance on the 400-record benchmark dataset, evaluation using independent clinical populations is still required to establish the model's external generalizability.



The experimental results indicate that the proposed hybrid framework can effectively extract discriminative information from heterogeneous CKD variables. Although the standalone CNN showed strong performance, combining its learned representations with conventional machine-learning classifiers provided further improvement. Among the evaluated models, CNN-XGBoost achieved 98.75% accuracy, 100% precision, 98% recall, 100% specificity, a 98.99% F1-score, a 0.9993 ROC-AUC, and a 0.9996 PR-AUC. The consistently strong performance of the other hybrid models also indicates that the learned CNN representation can be effectively used with different classifier types. However, validation on larger multicentre CKD cohorts and independent external datasets is necessary before the proposed model can be considered for clinical implementation or real-world diagnostic applications.

5. Model Explainability and Clinical Interpretation

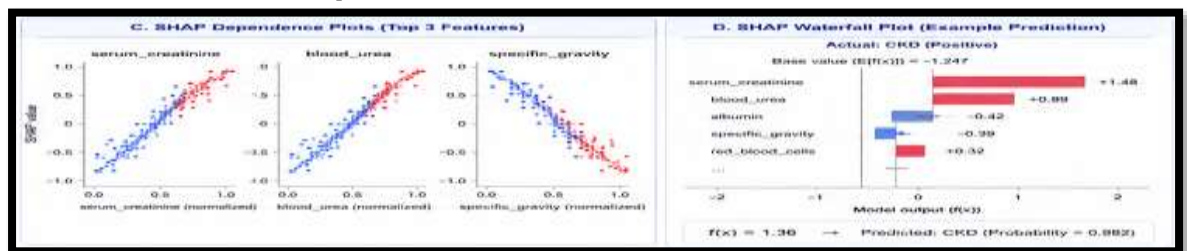
High predictive accuracy by itself does not establish clinical usefulness; interpretation also requires understanding which variables influence a prediction, the direction of their influence, and their contribution to individual cases. Accordingly, SHAP (SHapley Additive exPlanations) was incorporated into the CNN-XGBoost analysis. Global SHAP results describe overall feature influence, whereas local explanations indicate how individual patient characteristics affect a particular prediction (Lundberg & Lee, 2017)[14].

5.1 SHAP-Based Feature Importance and Clinical Interpretation: The SHAP summary and dependence analyses showed that the contribution of clinical variables varies across observations. The 64-dimensional CNN representation, when combined with XGBoost, therefore captured nonlinear patterns that could be related back to recognizable clinical variables through SHAP, improving interpretability of the final predictions.

Within the CNN-XGBoost model, serum creatinine had the largest overall SHAP importance (0.184), followed by blood urea (0.127), specific gravity (0.104), albumin (0.093), red blood cells (0.081), sugar (0.074), hemoglobin (0.063), blood pressure (0.059), age (0.057), and hypertension (0.048). Higher serum creatinine and blood urea generally pushed predictions toward CKD, whereas specific gravity showed an inverse contribution pattern.

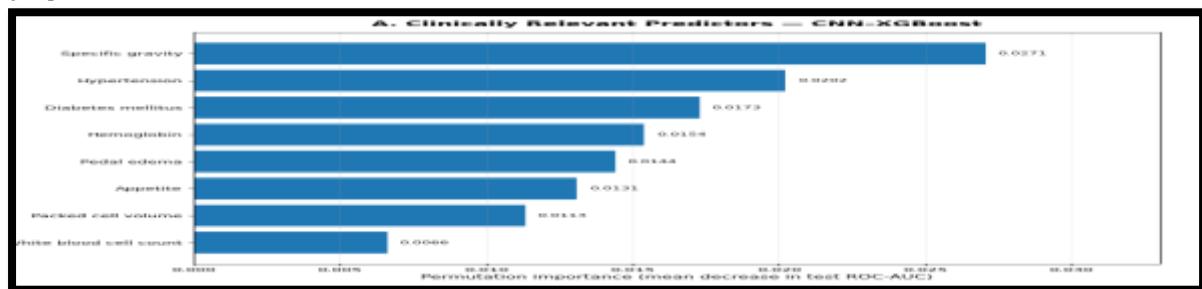


The largest positive SHAP contributions were observed for serum creatinine (+1.48), blood urea (+0.89), and red blood cells (+0.32), while albumin (−0.42) and specific gravity (−0.39) contributed in the opposite direction. Thus, SHAP provides case-level insight by connecting model outputs with clinically recognizable variables. These values should, however, be interpreted as model-attribution measures rather than evidence of causal relationships.



5.2 Clinically Relevant Predictors and Risk Stratification: The CNN-XGBoost model provided strong discrimination together with feature-importance information and probability-based risk stratification. Specific gravity had the greatest permutation importance (0.0271), while SHAP placed serum creatinine and blood urea at the forefront of prediction, followed by hemoglobin, specific gravity, albumin, packed cell volume, blood pressure, age, diabetes mellitus, and hypertension. Increased creatinine and blood urea were associated with higher CKD predictions, whereas lower hemoglobin contributed more strongly toward the CKD class. Together, these analyses offer complementary views of predictor importance and model-based risk assessment.

5.2.1 Clinically Relevant Predictors: Permutation importance ranked specific gravity (0.0271) first, followed by hypertension (0.0202), diabetes mellitus (0.0173), hemoglobin (0.0154), pedal edema (0.0144), appetite (0.0131), packed cell volume (0.0113), and white blood cell count (0.0066). The results indicate that renal and hematological measurements, together with clinically relevant comorbidities and symptoms, contributed to the model's discrimination of CKD cases.



5.2.2 CNN-XGBoost Prediction Performance: Using the independent 80-observation test set, CNN-XGBoost achieved 98.75% accuracy, 100.00% precision, 98.00% recall, 100.00% specificity, 98.99% F1-score, and 0.9993 ROC-AUC. Among the CNN-based alternatives, it delivered the highest accuracy, F1-score, and ROC-AUC, indicating the strongest discriminatory performance in the evaluated setting.

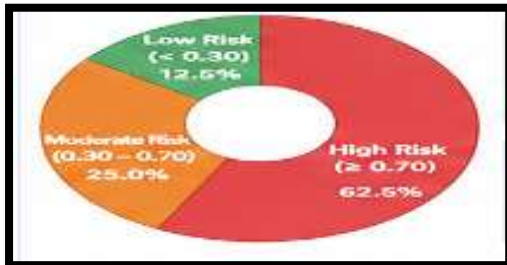
1. SUMMARY OF TEST SET PERFORMANCE (60 SAMPLES)

Model	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F1-Score (%)	ROC-AUC
CNN	97.50	98.00	98.00	98.67	98.00	0.9987
CNN-Random Forest	97.50	98.00	98.00	98.67	98.00	0.9988
CNN-XGBoost	98.75	100.00	98.00	100.00	98.99	0.9993
CNN-SVM	97.50	100.00	98.00	100.00	97.99	0.9981
CNN-Logistic Regression	97.50	100.00	98.00	100.00	97.99	0.9979

5.2.3 ROC-AUC and Precision-Recall Analysis: The ROC curves demonstrate excellent discrimination across all evaluated CNN-based models. CNN-XGBoost achieved the highest ROC-AUC (0.9993), followed by CNN (0.9987), CNN-Random Forest (0.9986), CNN-SVM (0.9981), and CNN-Logistic Regression (0.9979). The

corresponding PR analysis also shows the strongest performance for CNN-XGBoost, with a PR-AUC of 0.9996.

5.2.4 Risk Stratification and Prediction Insights: Based on predicted probabilities, the 400 observations were grouped into three risk categories: 50 low-risk cases (12.5%), 100 moderate-risk cases (25.0%), and 250 high-risk cases (62.5%). This probability-based grouping offers an interpretable way to prioritize cases for further assessment, although it should not be treated as a substitute for established clinical diagnosis.



5.3 Clinical Interpretation and Decision-Support Potential: The CNN-XGBoost framework combines strong predictive results with an interpretable representation of CKD-associated patterns. SHAP highlighted predictors spanning renal, biochemical, hematological, urinalysis, physiological, and comorbidity domains. These findings suggest possible value for CKD prediction and patient-level risk stratification as a decision-support aid. Because SHAP describes associations within the model rather than causation, the system should be used alongside clinical expertise and accepted diagnostic criteria. Prospective studies and external validation are required before implementation in clinical practice.

6. Discussion

This investigation developed a hybrid approach in which 1-D CNN feature learning was followed by RF, XGBoost, SVM, or LR classification. From the 34 processed clinical variables, the CNN generated a compact 64-dimensional representation. Across the tested configurations, CNN-XGBoost produced the most favourable combination of discrimination, accuracy, and specificity on the study dataset.

CNN-XGBoost was the leading model among the standalone CNN and hybrid alternatives. On the independent test set, it achieved 98.75% accuracy, 100.00% precision, 98.00% recall, 100.00% specificity, 98.99% F1-score, ROC-AUC of 0.9993, and PR-AUC of 0.9996.

MAJOR FINDINGS	
The proposed hybrid framework (CNN-XGBoost) achieved the best overall performance across all evaluation metrics.	
Highest Accuracy: 98.75% and F1-Score: 98.99% on the independent test set.	
Perfect Specificity (100%) with only 1 false negative and 0 false positives, minimizing misclassification risk.	
Best Discriminative Ability: ROC-AUC = 0.9993 and PR-AUC = 0.9996, outperforming all compared models.	
CNN-derived 64-D deep features captured complex, non-linear clinical patterns and enhanced downstream ML performance.	
Robust across 5-fold cross-validation and different imputation strategies, demonstrating stable and consistent performance.	

The CNN-XGBoost confusion matrix contained 49 true-positive cases, 30 true-negative cases, no false positives, and one false negative. This corresponds to strong separation of the two outcome classes on the independent test data.

Confusion Matrix of Best Hybrid Model (CNN-XGBoost)			
Actual	Predicted		
	CKD (Positive)	Not CKD (Negative)	
CKD (Positive)	49 (TP)	1 (FN)	Accuracy = 98.75% Precision = 100.00% Recall (Sensitivity) = 98.00% Specificity = 100.00% F1-Score = 98.99% ROC-AUC = 0.9993
Not CKD (Negative)	0 (FP)	30 (TN)	

Relative to the standalone CNN and the other hybrid configurations, CNN-XGBoost yielded higher accuracy and F1-score, indicating that XGBoost was able to exploit the nonlinear information encoded by the CNN representation effectively. Although the conventional RF, XGBoost, and SVM baselines also produced perfect scores on the reported test setting, the CNN-XGBoost model offered a competitive and balanced combination of deep feature learning and boosting-based classification.

6.2 Comparison with Previous Studies: The proposed model is compared with seven closely related studies in the CKD prediction domain. The comparison table presents the CNN-XGBoost model alongside previously reported approaches based on accuracy and AUC. However, differences in datasets, preprocessing, feature selection, and validation methods mean that the reported results should be interpreted as study-specific findings rather than direct head-to-head comparisons [15,17,19,22,24,29,38]. The proposed CNN-XGBoost model achieved 98.75% accuracy and an AUC of 0.9993. Although its accuracy is not the highest, its AUC exceeded those reported in the comparison, demonstrating strong discriminatory performance on the evaluated dataset. However, given the relatively small and widely reused 400-record

UCI CKD dataset, these results should not be considered evidence of broad clinical generalizability. Validation using larger and more diverse clinical datasets is therefore required.

S.N.	Author(s)	Year	Research Title	Dataset	Model	Accuracy	AUC
1	Polat et al.	2017	Diagnosis of Chronic Kidney Disease Based on Support Vector Machine by Feature Selection Methods	UCI CKD (400)	SVM	98.50%	0.988
2	Ali et al.	2020	Cost-Sensitive Ensemble Feature Ranking and Automatic Threshold Selection for Chronic Kidney Disease Diagnosis	UCI CKD (400)	Cost-sensitive ensemble/RF	96.70%	0.982
3	Rashed-Al-Mahfuz et al.	2021	Clinically Applicable Machine Learning Approaches to Identify Attributes of Chronic Kidney Disease (CKD) for Use in Low-Cost Diagnostic Screening	UCI CKD (400)	Random Forest	97.00%	0.970
4	Ehsanedoh-Mtenye et al.	2022	A Machine Learning Method with Filter-Based Feature Selection for Improved Prediction of Chronic Kidney Disease	UCI CKD (400)	XGBoost	98.90%	0.990
5	Ganie et al.	2023	Chronic Kidney Disease Prediction Using Boosting Techniques Based on Clinical Parameters	UCI CKD (400)	AdaBoost	98.47%	0.986
6	Ghosh & Khundoker	2024	Investigation on explainable machine learning models to predict chronic kidney diseases	Clinical CKD dataset (491)	XGBoost	93.29%	0.9689
7	Al-Rashed et al.	2025	Classifying kidney disease using a dense layers deep learning model	Public CKD dataset (400)	Dense-layer deep learning model	99.00%	0.9821
8	Proposed Study	2026	Hybrid CNN-XGBoost for Chronic Kidney Disease Classification	UCI CKD (400)	Hybrid CNN-XGBoost	98.75%	0.9993

6.3 Framework Performance, Strengths, and Limitations: The CNN-XGBoost framework shows promise for CKD classification using routinely collected clinical variables, with 98.75% accuracy, 98% recall, 100% specificity, 0.9993 ROC-AUC, and 0.9996 PR-AUC. Its design combines CNN-based representation learning with several ML classifiers and evaluates them using cross-validation and multiple performance measures. Important limitations include the modest sample size, use of a single dataset, class imbalance, retrospective nature of the data, and the remaining limits of model interpretability and external validity. The framework should therefore be regarded as a proof of concept rather than a clinically deployed diagnostic system. Independent, multicentre, prospective datasets are required to examine generalizability, calibration, fairness, and clinical utility before real-world use.

7. Conclusion

This study has developed and evaluated an explainable hybrid CNN-machine learning framework for CKD classification by combining 1-D CNN-based representation learning with Random Forest, XGBoost, SVM, and Logistic Regression classifiers. Among the evaluated configurations, CNN-XGBoost has demonstrated the strongest performance on the independent test set, achieving 98.75% accuracy, 100% precision, 98% recall, 100% specificity, a 98.99% F1-score, a 0.9993 ROC-AUC, and a 0.9996 PR-AUC. SHAP analysis has further identified serum creatinine and blood urea as important contributors to model predictions. Overall, the findings indicate that combining CNN-derived feature representations with XGBoost provides a potentially effective and interpretable approach to CKD classification using structured clinical data. However, the findings are limited by the use of a single-source dataset comprising 400 observations. Therefore, the reported performance does not establish the clinical effectiveness or generalizability of the proposed framework beyond the evaluated dataset. Further studies should validate the model using larger, independent, multicentre, and prospective clinical cohorts and should assess its generalizability, calibration, fairness, and clinical utility. Such validation will be essential before the framework can be considered for integration into real-world clinical decision-support systems.

Acknowledgements

The author acknowledges Amity University Uttar Pradesh, Greater Noida Campus, for its institutional support and research facilities, and the UCI Machine Learning Repository for providing the CKD dataset. Gratitude is also extended to the research community, family, academic mentors, colleagues, and others for their support, encouragement, and valuable contributions throughout the research.

Conflict of Interest

The author declares no financial, personal, professional, or institutional conflicts of interest that could have influenced this research or its publication.

References

- [1].Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20, 273–297. <https://doi.org/10.1007/BF00994018>
- [2].Breiman, L. (2001). Random forests. Machine Learning, 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
- [3].Fawcett, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters, 27(8), 861–874. <https://doi.org/10.1016/j.patrec.2005.10.010>
- [4].Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted Boltzmann machines. Proceedings of the 27th International Conference on Machine Learning, 807–814.

- [5].Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- [6].Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). John Wiley & Sons. <https://doi.org/10.1002/9781118548387>
- [7].Lin, M., Chen, Q., & Yan, S. (2013). Network in network. *arXiv*.
- [8].Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- [9].LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521, 436–444. <https://doi.org/10.1038/nature14539>
- [10].Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *Proceedings of the 32nd International Conference on Machine Learning*, 37, 448–456.
- [11].Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS ONE*, 10(3), e0118432. <https://doi.org/10.1371/journal.pone.0118432>
- [12].UCI Machine Learning Repository. (2015). Chronic Kidney Disease dataset. University of California, Irvine.
- [13].Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. <https://doi.org/10.1145/2939672.2939785>
- [14].Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30.
- [15].Polat, H., Mehr, H. D., & Cetin, A. (2017). Diagnosis of chronic kidney disease based on support vector machine by feature selection methods. *Journal of Medical Systems*, 41(4), 55. <https://doi.org/10.1007/s10916-017-0703-x>
- [16].Almansour, N. A., Syed, H. F., Khayat, N. R., Altheeb, R. K., Juri, R. E., Alhiyafi, J., Alrashed, S., & Olatunji, S. O. (2019). Neural network and support vector machine for the prediction of chronic kidney disease: A comparative study. *Computers in Biology and Medicine*, 109, 101–111. <https://doi.org/10.1016/j.combiomed.2019.04.017>
- [17].Ali, S. I., Ali, B., Hussain, J., Hussain, M., Satti, F. A., Park, G. H., & Lee, S. (2020). Cost-sensitive ensemble feature ranking and automatic threshold selection for chronic kidney disease diagnosis. *Applied Sciences*, 10(16), 5663. <https://doi.org/10.3390/app10165663>
- [18].Ma, F., Sun, T., Liu, L., & Jing, H. (2020). Detection and diagnosis of chronic kidney disease using deep learning-based heterogeneous modified artificial neural network. *Future Generation Computer Systems*, 111, 17–26. <https://doi.org/10.1016/j.future.2020.04.036>
- [19].Rashed-Al-Mahfuz, M., Haque, A., Azad, A., Alyami, S. A., Quinn, J. M. W., & Moni, M. A. (2021). Clinically applicable machine learning approaches to identify attributes of chronic kidney disease (CKD) for use in Low-Cost diagnostic screening. *IEEE Journal of Translational Engineering in Health and Medicine*, 9, 1–11. <https://doi.org/10.1109/jtehm.2021.3073629>
- [20].Senan, Ebrahim Mohammed, Al-Adhaileh, Mosleh Hmoud, Alsaade, Fawaz Waselallah, Aldhyani, Theyazn H. H., Alqarni, Ahmed Abdullah, Alsharif, Nizar, Uddin, M. Irfan, Alahmadi, Ahmed H., Jadhav, Mukti E, Alzahrani, Mohammed Y., Diagnosis of Chronic Kidney Disease Using Effective Classification Algorithms and Recursive Feature Elimination Techniques, *Journal of Healthcare Engineering*, 2021, 1004767, 10 pages, 2021. <https://doi.org/10.1155/2021/1004767>
- [21].Debal, D. A., & Sitote, T. M. (2022). Chronic kidney disease prediction using machine learning techniques. *Journal of Big Data*, 9, 109. <https://doi.org/10.1186/s40537-022-00657-5>
- [22].Ebiaredoh-Mienye, S. A., Swart, T. G., Ezenogho, E., & Mienye, I. D. (2022). A Machine Learning Method with Filter-Based Feature Selection for Improved Prediction of Chronic Kidney Disease. *Bioengineering*, 9(8), Article 350. <https://doi.org/10.3390/bioengineering9080350>
- [23].Loftus, T. J., Shickel, B., Ozrazgat-Baslanti, T., Ren, Y., Glicksberg, B. S., Cao, J., Singh, K., Chan, L., Nadkarni, G. N., & Bihorac, A. (2022). Artificial intelligence-enabled decision support in nephrology. *Nature Reviews Nephrology*, 18(7), 452–465. <https://doi.org/10.1038/s41581-022-00562-3>
- [24].Ganie, S. M., Dutta Pramanik, P. K., Mallik, S., & Zhao, Z. (2023). Chronic kidney disease prediction using boosting techniques based on clinical parameters. *PLOS ONE*, 18(12), e0295234. <https://doi.org/10.1371/journal.pone.0295234>
- [25].International Society of Nephrology. (2023). ISN–Global Kidney Health Atlas 2023. International Society of Nephrology.
- [26].Singh, D. P., Jassi, J. S., & Sunaina. (2023). Exploring the significance of statistics in research: A comprehensive overview. *European Chemical Bulletin*, 12(Special Issue 2), 2089–2102.

- [27].Sanmarchi, F., Fanconi, C., Golinelli, D., Gori, D., Hernandez-Boussard, T., & Capodici, A. (2023). Predict, diagnose, and treat chronic kidney disease with machine learning: A systematic literature review. *Journal of Nephrology*, 36(4), 1101–1117. <https://doi.org/10.1007/s40620-023-01573-4>
- [28].Chhabra, D., Juneja, M., & Chutani, G. (2024). An efficient ensemble-based machine learning approach for predicting chronic kidney disease. *Current Medical Imaging*, 20. <https://doi.org/10.2174/1573405620666230508104538>
- [29].Ghosh, S. K., & Khandoker, A. H. (2024). Investigation on explainable machine learning models to predict chronic kidney diseases. *Scientific Reports*, 14, 3687. <https://doi.org/10.1038/s41598-024-54375-4>
- [30].Kidney Disease: Improving Global Outcomes (KDIGO). (2024). KDIGO 2024 clinical practice guideline for the evaluation and management of chronic kidney disease. *Kidney International*, 105(4S), S117–S314.
- [31].Nagawa, K., Hara, Y., Inoue, K., Yamagishi, Y., Koyama, M., Shimizu, H., Matsuura, K., Osawa, I., Inoue, T., Okada, H., Kobayashi, N., & Kozawa, E. (2024). Three-dimensional convolutional neural network-based classification of chronic kidney disease severity using kidney MRI. *Scientific Reports*, 14, 15775. <https://doi.org/10.1038/s41598-024-66814-3>
- [32].Saif, D., Sarhan, A. M., & Elshennawy, N. M. (2024). Deep-kidney: An effective deep learning framework for chronic kidney disease prediction. *Health Information Science and Systems*, 12, 3. <https://doi.org/10.1007/s13755-023-00261-8>
- [33].Singh, D. P. (2024). An extensive analysis of machine learning models to predict breast cancer recurrence. *Tuijin Jishu/Journal of Propulsion Technology*, 45(2).
- [34].Singh, D. P. (2024). An extensive analysis of machine learning techniques for predicting the onset of lung cancer. *Tuijin Jishu/Journal of Propulsion Technology*, 45(4).
- [35].Singh, D. P. (2024). An extensive examination of machine learning methods for identifying diabetes. *Tuijin Jishu/Journal of Propulsion Technology*, 45(2).
- [36].Singh, D. P. (2024). Comprehensive analysis of machine learning models for cardiovascular disease detection and diagnosis. *Tuijin Jishu/Journal of Propulsion Technology*, 45(4).
- [37].Zheng, J.-X., et al. (2024). Interpretable machine learning for predicting chronic kidney disease progression risk. *DIGITAL HEALTH*. <https://doi.org/10.1177/20552076231224225>
- [38].Al-Rasheed, A., Saqib, S. M., Asghar, M. Z., Mazhar, T., Osman, A. S. A., Shahid, M., Iqbal, M., & Khan, M. A. (2025). Classifying kidney disease using a dense layers deep learning model. *SLAS Technology*, 33, 100324. <https://doi.org/10.1016/j.slant.2025.100324>
- [39].Gogoi, P., & Valan, J. A. (2025). Machine learning approaches for predicting and diagnosing chronic kidney disease: Current trends, challenges, solutions, and future directions. *International Urology and Nephrology*, 57(4), 1245–1268. <https://doi.org/10.1007/s11255-024-04281-5>
- [40].Ramu, K., Patthi, S., Prajapati, Y. N., Ramesh, J. V. N., Banerjee, S., Brahma Rao, K. B. V., Alzahrani, S. I., & Ayyasamy, R. (2025). Hybrid CNN-SVM model for enhanced early detection of chronic kidney disease. *Biomedical Signal Processing and Control*, 100, 107084. <https://doi.org/10.1016/j.bspc.2024.107084>
- [41].Singh, D. P. (2025). A comparative analysis of machine learning techniques for hypertension risk prediction and diagnostic classification. *International Journal of Innovative Research in Technology*, 11(12), 7183–7195.
- [42].Singh, D. P. (2025). Performance evaluation of advanced machine learning models for chronic kidney disease classification and prediction. *Tuijin Jishu/Journal of Propulsion Technology*, 46(3), 554–570. <https://propulsiontechjournal.com/index.php/journal/index>
- [43].Tangri, N., & Sabanayagam, C. (2025). Artificial intelligence approaches to enable early detection of CKD. *Nature Reviews Nephrology*, 21(3), 153–154. <https://doi.org/10.1038/s41581-025-00933-6>