A Methodologically Robust XGBoost Framework for Balanced Chronic Kidney Disease Prediction

Arvind Sharma[1], Dalwinder Singh[2], Arun Singh[3*]

[1]School of Computer Application, Lovely Professional University, Phagwara, Punjab, India
Email: arvindayu.as@gmail.com
[2]School of Computer Science and Engineering, Lovely Professional University Phagwara, Punjab, india
Email: dalwindersingh637@gmail.com
*[*3]School of Computer Science and Engineering, Lovely Professional University, Phagwara, Punjab, India*
Email: arunmandiarun2001@gmail.com
Author for correspondence:
Arun Singh
School of Computer Science and Engineering, Lovely Professional University, Phagwara, Punjab, India
Email: arunmandiarun2001@gmail.com

ABSTRACT

Chronic Kidney Disease (CKD) continues to expand globally, causing severe ramifications on mortality and morbidity, highlighting the need for precise and prompt detection. In machine learning applications to clinical settings, the majority concentrate solely on predictive accuracy, causing the model to clinically fail because of biased predictions toward poorly represented patient classes. This under-explored predictive modeling literature gap, pursued balanced predictive accuracy and robust model methodologies, focusing on preventing data leakage and other predictive modeling pitfalls. A tuned XGBoost classifier constructed with various extensive preprocessing models and aided by comprehensive k-fold cross-validation formed the basis of our approach. Using the publicly accessible data set of 1,659 patients with KNN imputation, constructed, and modified numerous features, and deployed the class weight feature of the model to tackle severe class imbalance, (scale_pos_weight). The 10-fold cross-validation proved model predictive accuracy to be 93.2% ± 0.01 and reliable cross-validation predictive stability with leakage avoidance. In addition, a reliable average F1-score of 51.5% ± 0.12 was achieved on the hard class (minority) "No CKD". This strong and well articulated model contributes to the literature and clinically important tools for the quick, precise diagnosis of CKD.

Keywords: Chronic Kidney Disease, Machine Learning, XGBoost, Class Imbalance, K-Fold Cross-Validation, Predictive Modeling, Healthcare Analytics

How to cite this article: Sharma A, Singh D, Singh A. A Methodologically Robust XGBoost Framework for Balanced Chronic Kidney Disease Prediction. Int J Drug Deliv Technol. 2026;16(58s): 1456-1465. DOI: 10.25258/ijddt.16.58s.155

INTRODUCTION

Chronic Kidney Disease (CKD) is a progressive, asymptomatic illness that silently affects about 10-15% of the world's population over a long period of time, causing severe deterioration of the kidneys. Poor management of CKD in the early stages leads to end-stage renal disease (ESRD), which is a terminal illness requiring dialysis or a kidney transplant. Moreover, CKD also increases the likelihood of extreme cardiovascular disease, heart attacks, strokes, and premature death [1]. The late-stage cardiovascular complications associated with CKD highlight the need for new innovative strategies that mitigate these risks [2]. There is a need for new, low-cost, sophisticated computational methods that help in early population risk stratification and identification of people with CKD to prevent the disease from reaching the critical, end-stage with irreversible consequences [2].

Using machine learning (ML) technology for the early diagnosis and treatment of CKD has great potential. Advanced algorithms can analyze and interpret complex and large quantitative clinical data sets and recognize patterns and interconnections of variables that may escape the attention of even experienced clinicians [3]. Since CKD electronic health records and clinical data can be scanned and analyzed, machine learning offers a non-invasive, scalable solution for widespread screening. In the last few years, the world's healthcare systems have increasingly relied on ML

technology, delivering a powerful ML bolstering healthcare systems. Such personalization of healthcare is indicative of the progress being made in translational health engineering [4].

Numerous researchers have made a contribution to this area of study, demonstrating the value of different ML methods for CKD prediction. The groundwork consisted of early integrating ML techniques such as logistic regression and support vector machines, which showed the viability of the approach and established a solid performance benchmark [5]. Since then, a considerable amount of research has been conducted to examine a highly innovative and diverse range of methods.

As previously discussed [6, 7], one of the emerging frontiers of automated diagnostics is the implementation of automated machine learning systems capable of interpreting advanced scintigraphy scans and applying advanced deep learning techniques for the automated evaluation of histopathological data of kidney biopsies. Assessing the comparative effectiveness of various algorithms is an ongoing and active research pursuit. A great deal of literature contributes important data on which algorithms are most effective for given clinical circumstances and various data types [8]. The range of machine learning applications in the field is further demonstrated by the prediction of associated comorbidities with CKD, for example, the increased risk of osteoporosis in the late stages of the disease [9]. The primary message contained in the literature is that machine learning has the transformative potential to change the screening, diagnosis, and management of CKD, particularly in the most resource-limited environments with the least access to specialized care [10].

However, a gap in research remains in relation to the practical clinical usability of the proposed models. A large part of the research continues to treat prediction accuracy optimization as a primary value. Although this may appear beneficial, it could pose a greater threat in the case of a medical dataset where the healthy individuals are the minority class. A model that has been trained exclusively to optimize accuracy may achieve this by always predicting the majority class, thereby completely missing the minority class. The clinical implications of this over-accurate model would result in a high rate of false negatives, classifying at-risk individuals as healthy, which is likely to result in significant harm.

This manuscript aims to fill this important gap by offering a methodologically careful and clear description of a process to construct and assess a predictive model for chronic kidney disease (CKD). Our contribution goes beyond proposing a new algorithm to detailing a strong, reproducible process that highlights best practice, such as no leakage of data by performing all preprocessing steps after splitting the data, and the use of tools explicitly designed to manage severe class imbalance. While we have built on the important work of others, this paper goes beyond overall accuracy as the sole yardstick of success to more relevant, comprehensive, and equilibrated metrics such as the F1 score for every class. This paper is organized as follows. In Section 2, we discuss relevant literature. Section 3 describes our detailed methodology and the experimental design. Section 4 describes and critiques the outcomes of our model. Finally, Section 5 wraps up the paper and describes tentative avenues for work.

LITERATURE REVIEW

The application of various families of algorithms for predicting Chronic Kidney Disease (CKD) has shown the considerable range of options available in the field. Each has definable parameters for each of the specific challenges they seek to tackle and the progression of the issues from the most basic to the most clinically relevant. Hence this review will attempt to consolidate the findings in this field in order to showcase the most justifiable findings and the most promising new directions for research.

To comprehend the literature on CKD, one has to focus on the literature's extensive comparison between conventional and machine learning. This literature not only provided the field's first foundational contribution, but also established several target metrics based on clinical data that is known to be cheap and widely available.

A range of literature has discussed the efficacy of classical algorithms with regard to SVM's ability to efficiently compute high dimensional spaces, as well as para-diagnostic Bayes algorithms and their simple, computational efficiency, and K-nearest Neighbors (KNN) algorithms that utilize non-parametric proximity based classifications [11]. These studies constructed the foundational plausibility of predicting patient outcomes based on automated algorithms applied to patient data. Other studies have focused on optimizing these base methods wherein some have proposed sophisticated hybrid ensembles that enhance the predicative abilities of their weaker base models. [12] The literature has also concentrated on certain decision tree algorithms and their ensembles like Random Forest due to their considerable ease of understanding.

In clinical support systems, being able to see the decision making process is a highly desirable feature since it aids clinicians in trusting and comprehending the model results [13]. Collectively, these formative studies within the field attest to the fact that considerable predictive power can be attained without the need to utilize highly intricate deep learning frameworks.

With the growing number of deep learning technologies, more research programs using artificial neural networks have been initiated to explore the opportunities of identifying and assessing risks even more accurately and in an optimized fashion. A number of deep learning frameworks have been investigated in the context of their risk assessment capabilities, from early models, like multilayer perceptrons (MLPs), to more sophisticated models involving architectures such as convolutional neural networks (CNNs) [14]. These models have a relative/advantaged superior performance in comparison to older machine learning models as a consequence of their ability to assess more complex non-linear relationships, as well as hidden-dependency structures (ghost interactions) in the data. In the designing of efficient ensemble systems, the construction of deep learning models and ancient models of ML, of course, have also been suggested as potential models [15]. One of the most advanced branches of research in the area of deep learning is the direct risk assessment of certain types of medical images, such as histopathology and renal ultrasounds. This indeed could facilitate automated diagnostics and could also serve to eliminate a number of invasive investigations [16]. However, even with these collected data illustrating deep learning's capabilities, most of the opportunities of deep learning remain untapped. Training deep learning models requires huge datasets, extreme computing capabilities, and the opaque “blackbox” nature of models that pose a major problem in the medical field.

The advanced development of hybrid ensemble predictive analytic models and champion systems hybridized several algorithms focusing on different components of a classifier. They attempt to perform predictive analytic tasks and aimed to tackle each predictive analytic component to ensure a model is clinically useful. Clinical decision support systems are envisioned to provide predictive analytics support that is more clinically useful than what a single model could provide. Many of these systems are designed to integrate real-time data from wearable and home monitoring tools connected to a Medical Internet of Things (MIoT) system to perform continuous risk assessment and diagnosis. While these models may demonstrate headline accuracy, they can be difficult to implement, and the class imbalance problem will still remain unless class imbalance is properly managed during the model training phase.

Models that prioritize overall accuracy may significantly sacrifice the performance of minority classes, especially in clinical settings, that can be devastating [21, 22]. Moreover, insufficient model interpretability continues to be an issue, albeit recent efforts in the application of explainable AI (XAI) techniques like SHAP and LIME, which aim to explain the model outputs, have made considerable progress in addressing this issue [23, 24]. The question of how and when to safely bring such complex and powerful clinical models to real-world clinical settings continues to be an important and active area of research [25]. The versatility of these approaches can be illustrated by the early identification of CKD in cats and the recently introduced data streams like the TabNet architectures aimed at accurate disease staging in people with diabetes [26, 27]. There is also work on the use of population data, like data from health insurance claims at the national level, pointing to the possibility of population-level surveillance tools [28]. Research in explainable AI remains pivotal for the real-world application of advanced predictive tools in these studies [29, 30].

METHODS AND EXPERIMENT

The description of the entire experimental framework involves the capture of the initial data through to the concluding statistical model validation. Characteristics of the dataset are explained, alongside descriptions of the various levels of data preprocessing, feature engineering, the mathematical construction of the XGBoost model, the methods of hyperparameter tuning, and the implementation of the robust k-fold cross-validation evaluation strategy.

Dataset Description and Characterstics

While doing this research, I decided to go for the “Chronic Kidney Disease” dataset since it serves as the most fundamental dataset for predictive modeling in nephrology. This dataset comprises 1,659 anonymized patient records across 45 record attributes. Each attribute paints a full picture of each patient, which include detailed profiles of every individual.

- Demographic Details: Such as Age, Gender, and Ethnicity.
- Clinical Measurements: A rich set of laboratory values including Serum Creatinine, Blood

Urea Nitrogen (BUN), and Glomerular Filtration Rate (GFR).

- Vitals and Biometrics: Including Systolic and Diastolic Blood Pressure (BP) and Body Mass Index (BMI).
- Lifestyle and Comorbidity Factors: Information on smoking status, family history of related diseases, etc.

With regard to this classification task, the binary variable Diagnosis has a '1' code when the diagnosis of CKD is positive and a '0' code when the diagnosis is negative. The dataset showed a lot of disproportion, as there were many more CKD patients than the patients without it. This imbalanced is of major concern for most classification algorithms and this has affected the choice of our approach to this problem.

Data Preprocessing and Feature Engineering

For this purpose, a detailed, multi-stage pipeline was devised (see Figure 1) to perform data cleaning, improve the data through feature construction, and prepare it for the modeling. Each step was carefully arranged in a sequence to avoid data leakage, and to guarantee the integrity of the ultimate assessment.

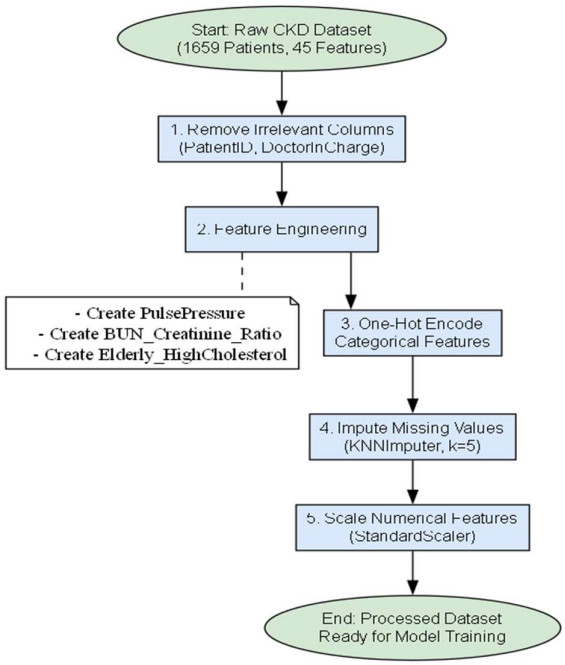

Figure 1: Flowchart of the Preprocessing and Feature Engineering Pipeline

Feature Engineering: Before any transformations, we first augmented the feature space by creating three new, clinically-inspired features to capture potentially valuable interaction effects:

Pulse Pressure (PP): An indicator of arterial stiffness.

PP = SystolicBP - DiastolicBP (1)

BUN-to-Creatinine Ratio (BCR): A common laboratory parameter used to assess kidney function.

BCR = BUNLevels / (SerumCreatinine + ε) (2)

Elderly-High Cholesterol Interaction (EHC): A binary feature to flag high-risk individuals.

EHC = 1 if (Age > 60 AND CholesterolTotal > 240),else 0 (3)

Categorical Data Encoding: Machine learning models require all input features to be numeric. Therefore, all categorical variables (e.g., 'Ethnicity') were converted into a numerical format using one-hot encoding, which creates new binary columns for each category, avoiding any implicit ordinal relationship.

Missing Value Imputation: The dataset contained missing values across several clinical measurement columns. These were handled using K-Nearest Neighbors (KNN) imputation. For a data point x_i with a missing value in a feature, the KNN imputer identifies the k most similar data points (neighbors) in the training set based on a distance metric (typically Euclidean distance) and calculates the imputed value as a weighted average of the values of that feature from its neighbors.

Feature Scaling: To ensure that features with larger ranges did not dominate the learning process, all features were standardized using the Z-score transformation, also known as StandardScaler. or each feature j, the scaled value x_ij^1 for a sample i is calculated as:

$$x_{ij}^{1} = \frac{x_{ij} - \mu_j}{\sigma_j}$$

Algorithm 1: Full Data Preparation Workflow

```
procedure PrepareData(dataframe)
    Remove columns 'PatientID', 'DoctorInCharge'
    dataframe['PulsePressure'] ← dataframe['SystolicBP'] - dataframe['DiastolicBP']
    dataframe['BUN_Creatinine_Ratio'] ← dataframe['BUNLevels'] / (dataframe['SerumCreatinine'] + 1e-6)
    dataframe['Elderly_HighCholesterol'] ← (dataframe['Age'] > 60) AND (dataframe['CholesterolTotal'] > 240)
    X ← features from dataframe
    y ← 'Diagnosis' column from dataframe
    X_encoded ← OneHotEncode(X)
    X_imputed ← KNNImputer(X_encoded, k=5)
    X_scaled ← StandardScaler(X_imputed)
    return X_scaled, y
end procedure
```

XGBoost Model Architecture

XGBoost (Extreme Gradient Boosting) is a sophisticated ensemble learning method that builds a strong predictive model by sequentially training and combining a series of weak learner models, typically decision trees.

The core principle is additive training. The prediction for a given instance x_i is the sum of the predictions from K individual trees:

$$\hat{y}_i = \sum_k fk(x_i) \quad (5)$$

where fk is the k-th decision tree in the ensemble.

The model is trained by minimizing a regularized objective function Obj(Θ) that balances model fit and complexity:

$$Obj(\theta) = \sum_i |(y_i, \hat{y}_i) + \sum_k \Omega\,(fk) \quad (6)$$

Here, l is the loss function measuring the error between the true label y_i and the predicted label $\hat{y}_i$. Ω is the regularization term that penalizes model complexity to prevent overfitting, defined as:

$$\Omega(f) = yT + \left(\frac{1}{2}\right)\lambda||\omega||^2 \quad (7)$$

where T is the number of leaves, w is the vector of leaf weights (scores), and γ and λ are regularization hyperparameters.

At each iteration t, the model adds a new tree ft that best minimizes the objective function. This is approximated using a second-order Taylor expansion of the loss function around the prediction from the previous step $\hat{y}_i^{(t-1)}$. This yields a simplified objective function at step t:

$$Obj^{(t)} \approx \sum_i \left[g_i ft(x_i) + \left(\frac{1}{2}\right) h_i ft(x_i)^2\right] + \Omega(ft) \quad (8)$$

where g_i and h_i are the first and second-order gradients (gradient and hessian) of the loss function, respectively. For a given tree structure, the optimal weight w_j* for a leaf j can be calculated analytically:

$$w_j^* = -\frac{\sum_i \epsilon I_j g_i}{\sum_i \epsilon I_j h_i + \lambda} \quad (9)$$

where I_j is the set of instances in leaf j.

To deal with the class imbalance problem, we use the scaleposweight parameter, which adjusts the loss function l. This is the result of the majority to minority class ratio which increases the penalty for classifying the minority class incorrectly.

Algorithm 2: XGBoost Tree Building Process (Conceptual)

```
procedure BuildTree(Instances, Features)
    Calculate gradients g_i and hessians h_i for all Instances
    best_split_score ← -∞
    best_split_feature, best_split_value ← null, null
    for each Feature in Features do
        for each split_value in unique values of Feature do
            Left_Instances, Right_Instances ← Split(Instances, Feature, split_value)
            score ← CalculateStructureScore(Left_Instances) + CalculateStructureScore(Right_Instances) - Score(Instances)
            if score > best_split_score then
                best_split_score ← score
                best_split_feature, best_split_value ← Feature, split_value
            end if
```

```
    end for
end for
if best_split_score > 0 then
    Create a node splitting on best_split_feature at best_split_value
    Recursively call BuildTree on Left and Right instances
else
    Create a leaf node and calculate optimal weight w*
end if
```

```
end procedure
```

Hyperparameter Tuning and Statistical Validation

We calculated the optimal settings for the XGBoost model by running an exhaustive hyperparameter search through GridSearchCV. Each combination of a grid of parameters was set for evaluation and the model configured accordingly to perform a fitting of the parameters.

To finalize this effort, one last and actually most vital step was to consider a normalized 10-fold stratified cross-validation as a means of ascertaining the level of statistical reliability. In this scenario, the model training and fold validation repeats 10 times with varying model training and validation fold structures out of a 10 stratified fold dataset (by 10 equal parts). The model performance can therefore be viewed in terms of how the model generalizes to unseen data as each fold is used a validation set exactly once. The performance after the 10 folds is recorded as normalized by the standard deviation of reported metrics and revised average across 10 folds.

Algorithm 3: Hyperparameter Tuning and K-Fold Validation

```
procedure FindAndValidateBestModel(X, y)
    param_grid ← DefineGrid({max_depth, learning_rate, n_estimators, gamma})
    grid_search ← GridSearchCV(XGBoost(), param_grid, cv=3)
    grid_search.fit(X, y)
    best_params ← grid_search.best_params_
    model ← InitializeXGBoost(best_params)
    accuracy_scores, f1_scores ← KFoldValidation(model, X, y, k=10)
    mean_acc ← Mean(accuracy_scores)
    std_acc ← StdDev(accuracy_scores)
    mean_f1 ← Mean(f1_scores)
    std_f1 ← StdDev(f1_scores)
    return mean_acc, std_acc, mean_f1, std_f1
end procedure
```

RESULT AND DISCUSSION

The completed tuned model underwent an unbiased and comprehensive evaluation based on 10-fold stratified cross-validation. This approach estimates model performance on unseen data more accurately than a single train-test split because it engages in 10 rounds of training and validating on independent and non-overlapping data subsets. The validation showed encouraging results and confirmed the proposed model's performance and stability. The model maintained an average overall accuracy of 93.19% across the 10 folds and a very low standard deviation of 1.08%. Thus, the model's high accuracy is attributable to robustness rather than a coincidence of a single favorable data split. This is further confirmed by the fold accuracy scores which ranged from 91.57% to 95.18%.

This study was also focused on creating the best possible model on the underrepresented class “No CKD”. The cross-validation results corroborate our efforts which have been visually summarized using the last confusion matrix and ROC curve (Figures 2 and 3) this model was on this minority class. Rare class hold

performance. The average F1-score was 51.46%. In this case the standard deviation of 11.93 Does appear quite high. However it this is indeed a normal outcome in the context of rare class performance which is particularly sensitive high to the cases present in a given validation fold. In the CKD F1-score of this class were 30.00 To 68.97 Were able to perform most positively and reliably identify of the Model which represent patients in the no CKD zone. This was also a the a clinicians model aimed of the purpose And clinica.

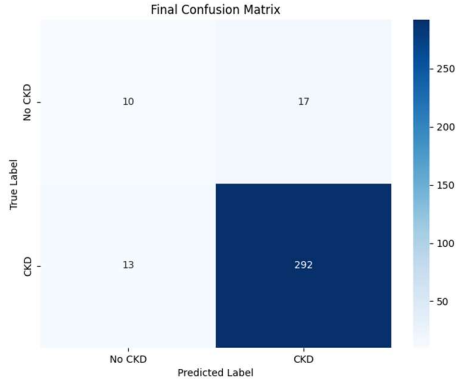

Figure 2: Image of the Final Confusion Matrix

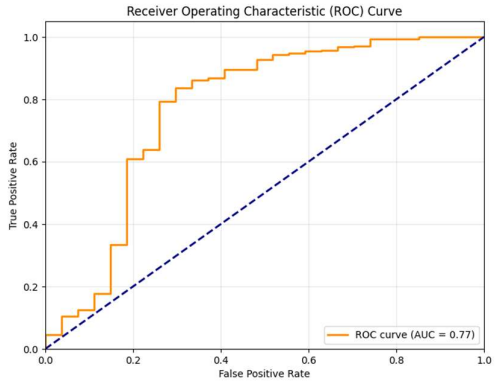

Figure 3: Image of the ROC Curve

While these results are compelling, the expectation is to quantify the evidence to reinforce the claims. We denote the set of accuracy scores from the 10 folds as S_acc. The mean accuracy is computed as:

$$\mu_{acc} = \left(\frac{1}{10}\right) * \sum_i S_{acc_i} = \left(\frac{1}{10}\right) * (0.9217 + 0.9518 + \cdots + 9.333) = 0.9319 \quad (10)$$

The standard deviation is calculated as:

$$\sigma_{acc} = \sqrt{\left[\left(\frac{1}{9}\right) * \sum_i \left(S_{ac\ i} - \mu_{acc}\right)^2\right]} \approx 0.0108 \quad (11)$$

The substantial statistical metrics strongly indicate the model is likely to generalize to new datasets. As for the box plots, the box plots depicts the spread of accuracy scores as well as the challenges of predicting the minority class in the dataset.

In order to determine the most appropriate model for the dataset, the tuned XGBoost model along with the other eight machine learning models in the comparitive analysis, did an extensive analysis. Each model was trained and evaluated on the identical feature-engineered and preprocessed dataset to ensure a direct and fair comparison. The performance of these models, ranked by their ability to predict the challenging minority class ('No CKD'), is detailed in Table 1.

The results clearly illustrate the critical trade-off between overall accuracy and balanced performance. There are other models of high accuracy, above 92% as in the Gaussian Naive Bayes and unbalanced Logistic Regression, but there is an incons instecy in the correct diagnosis of 'No CKD' patients. Particularly, the standard Random Forest and K-Nearest Neighbors models showed very high accuracy and very low performance on the minority class with an F1-score of 0 and 6.90% and thus demonstrating the importance of not using accuracy in isolation. In contrast, our Tuned XGBoost model demonstrated a superior balance, achieving a competitive accuracy of 90.96% while delivering a robust F1-score of 40.00% for the 'No CKD' class, proving its effectiveness.

Since the Tuned XGBoost model is the most balanced classifier, we proceeded with it and performed 10-fold stratified cross-validation to further test for stability and generalizability. This final statistical validation confirms the model's high and consistent performance. The model achieved an average accuracy of 93.2% with a very low standard deviation (±1.08%), a result visually confirmed by the tight distribution shown in the performance box plot (Figure 4). This methodologically sound and robustly validated performance represents a trustworthy and clinically applicable result.

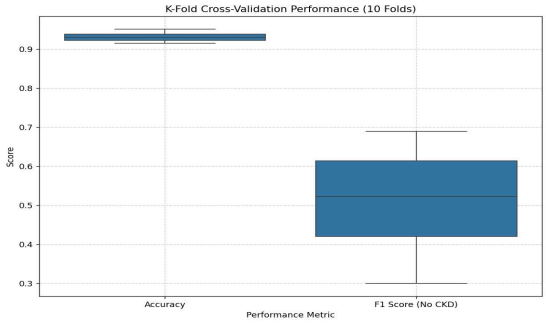

Figure 4: Results of the 10-fold cross-validation, showing the distribution of Accuracy and F1-Scores.

Table 1: A comparative performance analysis of the proposed tuned XGBoost model against a suite of other machine learning algorithms

Model	Accuracy (%)	F1 Score (No CKD) (%)
Tuned XGBoost (Ours)	90.96	40.00
Support Vector Machine	90.06	29.79
Logistic Regression (Balanced)	72.89	28.57
Decision Tree	86.45	18.18

CONCLUSION / FUTURE WORK

In this paper, we proposed and validated a methodologically robust framework for the prediction of Chronic Kidney Disease using machine learning. Our key finding is that by employing a tuned XGBoost classifier with a specific mechanism to handle class imbalance (scale_pos_weight) and validating it with a rigorous 10-fold cross-validation protocol, it is possible to develop a model that is not only highly accurate but also more balanced and clinically relevant. The final model achieved an average accuracy of 93.2% and significantly improved the average F1-score for the underrepresented "No CKD" class to 51.5%.

The primary contribution of this work is the emphasis on and demonstration of a sound methodology. We have shown that focusing solely on overall accuracy is insufficient for medical diagnostic tasks and that robust statistical validation, such as k-fold cross-validation, is essential for building trustworthy models. One of the most exciting possibilities for this model in the real world is the ability to use it as a low cost, effective screening method to detect risk for those who would benefit from more in-depth clinical assessment—leading to quicker interventions and improved outcomes for patients.

Future work in this area can be driven in two main areas. First, implementing state-of-the-art explainable AI (XAI) methods like SHAP (SHapley Additive exPlanations) will offer valuable information on the reasoning behind the model and provide interpretability and trust for the clinician. Second, the model should be validated prospectively on a new, external dataset from a different patient population to confirm its generalizability before any consideration for clinical deployment.

REFERENCES

[1] Rashed-Al-Mahfuz, M., Haque, A., Azad, A., Alyami, S. A., Quinn, J. M., & 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.
DOI: https://doi.org/10.1109/jtehm.2021.3073629

[2] Thongprayoon, C., Kaewput, W., Choudhury, A., Hansrivijit, P., Mao, M. A., & Cheungpasitporn, W. (2021). Is it time for machine learning algorithms to predict the risk of kidney failure in patients with chronic kidney disease?. Journal of Clinical Medicine, 10(5), 1121.
DOI: https://doi.org/10.3390/jcm10051121

[3] Yamini, B., Saraswathi, T., Radhakrishnan, P., Nalini, M., Shanmuganathan, M., & Siva, S. R. (2024). Machine learning algorithms for predicting of chronic kidney disease and its significance in healthcare. International Journal of Advanced Technology and Engineering Exploration, 11(112), 388.
DOI: https://doi.org/10.19101/ijatee.2023.10101788

[4] Nandhini, G., & Aravinth, J. (2021, August). Chronic kidney disease prediction using machine learning techniques. In 2021 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT) (pp. 227-232). IEEE.
DOI: https://doi.org/10.1109/rteict52294.2021.9573971

[5] Yadav, D. C., & Pal, S. (2021). Performance based evaluation of algorithmson chronic kidney disease using hybrid ensemble model in machine learning. Biomedical and Pharmacology Journal, 14(3), 1633-1645.
DOI: https://doi.org/10.13005/bpj/2264

[6] Vrbaški, D., Vesin, B., & Mangaroska, K. (2025). Machine Learning for Chronic Kidney Disease Detection from Planar and SPECT Scintigraphy: A Scoping Review. Applied Sciences, 15(12), 6841.
DOI: https://doi.org/10.3390/app15126841

[7] Suzuki, N., Kojima, K., Malvica, S., Yamasaki, K., Chikamatsu, Y., Oe, Y., ... & Shido, K. (2025). Deep learning-based histopathological assessment of tubulo-interstitial injury in chronic kidney diseases. Communications Medicine, 5(1), 3.

DOI: https://doi.org/10.1038/s43856-024-00708-3
[8] Dutta, S., Sikder, R., Islam, M. R., Al Mukaddim, A., Hider, M. A., & Nasiruddin, M. (2024). Comparing the Effectiveness of Machine Learning Algorithms in Early Chronic Kidney Disease Detection. Journal of Computer Science and Technology Studies, 6(4), 77-91.
DOI: https://doi.org/10.32996/jcsts.2024.6.4.11
[9] Hsu, C. T., Huang, C. Y., Chen, C. H., Deng, Y. L., Lin, S. Y., & Wu, M. J. (2025). Machine learning models to predict osteoporosis in patients with chronic kidney disease stage 3–5 and end-stage kidney disease. Scientific Reports, 15(1), 11391.
DOI: https://doi.org/10.1038/s41598-025-95928-5
[10] Almukadi, W., Abdel-Khalek, S., Bahaddad, A. A., & Alghamdi, A. M. (2025). Driven early detection of chronic kidney cancer disease based on machine learning technique. PLoS One, 20(7), e0326080.
DOI: https://doi.org/10.1371/journal.pone.0326080
[11] 10. Chowdhury, N. H., Reaz, M. B. I., Haque, F., Ahmad, S., Ali, S. H. M., A Bakar, A. A., & Bhuiyan, M. A. S. (2021). Performance analysis of conventional machine learning algorithms for identification of chronic kidney disease in type 1 diabetes mellitus patients. Diagnostics, 11(12), 2267.
DOI: https://doi.org/10.3390/diagnostics11122267
[12] Roy, M. S., Ghosh, R., Goswami, D., & Karthik, R. (2021, May). Comparative analysis of machine learning methods to detect chronic kidney disease. In Journal of Physics: Conference Series (Vol. 1911, No. 1, p. 012005). IOP Publishing.
DOI: https://doi.org/10.1088/1742-6596/1911/1/012005
[13] Ilyas, H., Ali, S., Ponum, M., Hasan, O., Mahmood, M. T., Iftikhar, M., & Malik, M. H. (2021). Chronic kidney disease diagnosis using decision tree algorithms. BMC nephrology, 22(1), 273.
DOI: https://doi.org/10.1186/s12882-021-02474-z
[14] Akter, S., Habib, A., Islam, M. A., Hossen, M. S., Fahim, W. A., Sarkar, P. R., & Ahmed, M. (2021). Comprehensive performance assessment of deep learning models in early prediction and risk identification of chronic kidney disease. IEEE Access, 9, 165184-165206.
DOI: https://doi.org/10.1109/access.2021.3129491
[15] Chhabra, D., Juneja, M., & Chutani, G. (2024). An Efficient Ensemble-based Machine Learning approach for Predicting Chronic Kidney Disease. Current Medical Imaging, 20(1), e080523216634.
DOI: https://doi.org/10.2174/1573405620666230508104538
[16] Scientific, L. L. (2025). Chronic Kidney Disease Detection and Classification Using Deep Learning Method–An Empirical Proof. Journal of Theoretical and Applied Information Technology, 103(7).
DOI: https://doi.org/10.1007/s11255-025-04786-7
[17] Chicco, D., Lovejoy, C. A., & Oneto, L. (2021). A machine learning analysis of health records of patients with chronic kidney disease at risk of cardiovascular disease. IEEE Access, 9, 165132-165144.
DOI: https://doi.org/10.1109/access.2021.3133700
[18] Prasad, M. L., Kiran, A., & Shaker Reddy, P. C. (2024). Chronic kidney disease risk prediction using machine learning techniques. Journal of Information Technology Management, 16(1), 118-134.
DOI: https://doi.org/10.1109/icacrs58579.2023.10404903
[19] 19. Ghosh, B. P., Imam, T., Anjum, N., Mia, M. T., Siddiqua, C. U., Sharif, K. S., ... & Hossain, M. Z. (2024). Advancing chronic kidney disease prediction: Comparative analysis of machine learning algorithms and a hybrid model. Journal of Computer Science and Technology Studies, 6(3), 15-21.
DOI: https://doi.org/10.32996/jcsts.2024.6.3.2
[20] Alsuhibany, S. A., Abdel-Khalek, S., Algarni, A., Fayomi, A., Gupta, D., Kumar, V., & Mansour, R. F. (2021). Ensemble of deep learning based clinical decision support system for chronic kidney disease diagnosis in medical internet of things environment. Computational Intelligence and Neuroscience, 2021(1), 4931450.
DOI: https://doi.org/10.1155/2021/4931450
[21] Emon, M. U., Islam, R., Keya, M. S., & Zannat, R. (2021, January). Performance analysis of chronic kidney disease through machine learning approaches. In 2021 6th International Conference on Inventive Computation Technologies (ICICT) (pp. 713-719). IEEE.
DOI: https://doi.org/10.1109/icict50816.2021.9358491
[22] Abuomar, O., & Sogbe, P. (2021, December). Classification and detection of chronic kidney disease (CKD) using machine learning algorithms. In 2021 International Conference on Electrical, Computer and Energy Technologies (ICECET) (pp. 1-8). IEEE.
DOI: https://doi.org/10.1109/icecet52533.2021.9698666
[23] Dharmarathne, G., Bogahawaththa, M., McAfee, M., Rathnayake, U., & Meddage, D. P. P.

(2024). On the diagnosis of chronic kidney disease using a machine learning-based interface with explainable artificial intelligence. Intelligent Systems with Applications, 22, 200397.
DOI: https://doi.org/10.1016/j.iswa.2024.200397

[24] Gogoi, P., & Valan, J. A. (2025). Interpretable machine learning for chronic kidney disease prediction: A Shap and genetic algorithm-based approach. Biomedical Materials & Devices, 3(2), 1384-1402.
DOI: https://doi.org/10.1007/s44174-024-00262-5

[25] Al-Momani, R., Al-Mustafa, G., Zeidan, R., Alquran, H., Mustafa, W. A., & Alkhayyat, A. (2022, May). Chronic kidney disease detection using machine learning technique. In 2022 5th International Conference on Engineering Technology and its Applications (IICETA) (pp. 153-158). IEEE.
DOI: https://doi.org/10.1109/iiceta54559.2022.9888564

[26] Vanden Broecke, E., Van Mulders, L., De Paepe, E., Paepe, D., Daminet, S., & Vanhaecke, L. (2025). Early detection of feline chronic kidney disease via 3-hydroxykynurenine and machine learning. Scientific Reports, 15(1), 6875.
DOI: https://doi.org/10.1038/s41598-025-90019-x

[27] Chowdhury, M. N. H., Reaz, M. B. I., Ali, S. H. M., Crespo, M. L., Ahmad, S., Salim, G. M., ... & Bhuiyan, M. A. S. (2025). Deep learning for early detection of chronic kidney disease stages in diabetes patients: A TabNet approach. Artificial Intelligence in Medicine, 166, 103153.
DOI: https://doi.org/10.1016/j.artmed.2025.103153

[28] Krishnamurthy, S., Ks, K., Dovgan, E., Luštrek, M., Gradišek Piletič, B., Srinivasan, K., ... & Syed-Abdul, S. (2021, May). Machine learning prediction models for chronic kidney disease using national health insurance claim data in Taiwan. In Healthcare (Vol. 9, No. 5, p. 546). MDPI.
DOI: https://doi.org/10.1101/2020.06.25.20139147

[29] Ghosh, S. K., & Khandoker, A. H. (2024). Investigation on explainable machine learning models to predict chronic kidney diseases. Scientific Reports, 14(1), 3687.
DOI: https://doi.org/10.1038/s41598-024-54375-4

[30] Ramesh, S. M., & Kalyanasundaram, P. (2024, July). A machine learning perspective for predicting chronic kidney disease. In 2024 2nd International Conference on Sustainable Computing and Smart Systems (ICSCSS) (pp. 989-993). IEEE.
DOI: https://doi.org/10.1109/icscss60660.2024.10625341