Boosting-Based Prediction of Chronic Kidney Disease Using Clinical Parameters

1. Lt S Babu, Research Scholar, Department of Computer Science, Karuppanan Mariappan College, Thirupur, Tamil Nadu, India. babusanjana@gmail.com

2.Dr. P. Parameswari , Principal, Palanisamy College of Arts, Perundurai, Erode, Tamil Nadu, India. paramtech20@gmail.com

Abstract

Chronic Kidney Disease (CKD) is a progressive condition that can lead to severe health complications if not diagnosed early. Accurate and timely prediction of CKD is crucial for effective treatment and management. This study explores the application of boosting techniques, such as Adaptive Boosting (AdaBoost), Gradient Boosting, and Extreme Gradient Boosting (XGBoost), to enhance the predictive accuracy of CKD using clinical parameters. A dataset comprising key medical indicators, including blood pressure, glucose levels, creatinine, and hemoglobin, was utilized for model training and evaluation. Feature selection techniques were employed to identify the most relevant parameters contributing to CKD prediction. The performance of boosting models was compared using accuracy, precision, recall, and F1-score metrics. Experimental results indicate that boosting-based models outperform traditional machine learning algorithms in terms of predictive accuracy and robustness. This paper highlights the significance of ensemble learning in medical diagnosis and suggests boosting techniques as reliable tools for CKD prediction. These findings can aid healthcare professionals in early detection and personalized treatment planning.

Keywords: ML, DL, CKD

1. INTRODUCTION

The kidneys are essential organs responsible for filtering waste, maintaining fluid balance, and activating vitamin D for bone health. They also regulate blood pressure by producing hormones such as renin. Impaired kidney function leads to fluid retention and increases the risk of conditions like heart disease, anemia, hypertension, and bone disease. Chronic Kidney Disease (CKD) is a major global health concern, with mortality and morbidity rates rising annually. According to the World Health Organization, CKD affects approximately 10% of the population, making it the eighth leading cause of death worldwide, claiming 1.2 million lives each year. Many CKD patients remain unaware of the disease in its early stages due to its asymptomatic nature, leading to late-stage diagnoses that require dialysis or kidney transplants. Regular medical checkups and blood tests are recommended for early detection.

CKD is classified into five stages based on the Glomerular Filtration Rate (GFR), which measures how effectively the kidneys filter waste. In later stages, kidney function deteriorates significantly, complicating treatment. Traditional diagnostic methods, including serum creatinine tests, blood urea nitrogen levels, and imaging techniques like X-rays, CT scans, and MRIs, often fail to provide early and precise detection. These methods also pose challenges such as radiation exposure and delayed diagnosis. Therefore, developing an advanced predictive model is crucial for the early identification of CKD.

Despite the existence of computerized prediction techniques, their accuracy remains suboptimal. This research utilizes a kidney disease dataset from Kaggle to improve early-stage CKD prediction using three machine learning models: Grey Wolf Optimized SVMBoost (GWO-SVMBoost), Quantum Particle Swarm Optimized Extreme Learning Machine with XGBoost (QPSO-ELMX), and FuzzyFireNet. To enhance model performance, the dataset undergoes preprocessing, and SMOTE is applied to address class imbalance issues. The models are evaluated based on accuracy, precision, recall, and training time. Among the three approaches, the FuzzyFireNet classifier demonstrates superior accuracy and faster training time, outperforming both QPSO-ELMX and GWO-SVMBoost.

2. LITERATURE SURVEY

Numerous research initiatives have explored the use of clinical and genetic data analysis for the detection and prediction of Chronic Kidney Disease (CKD). Salmon et al. [1] developed a hybrid model combining Random Forest and Logistic Regression, achieving a 95% accuracy rate—2% higher than existing models—though it struggled with high dimensionality. To reduce execution delay and improve classification, Lee et al. [2] introduced the Common Data Model, which achieved an 80% improvement in accuracy but still faced challenges related to high-dimensional data.

Othman [3] applied Logistic Regression with Support Vector Machines for CKD prediction, attaining an accuracy of 85% and sensitivity of 78%, though dimensionality remained a concern. Kulkarni et al. [4] utilized Random Forest and Logistic Regression to predict multiple chronic diseases, achieving a 78% accuracy rate for kidney disease. Korke et al. [5] designed a complex model with 92% accuracy, showing strong diagnostic capability while reducing errors. However, issues such as poor image quality, communication gaps, and physician workload hindered early detection, leading to increased medical costs and negative patient outcomes.

Ramani et al. [6] proposed a machine learning-based model with 79.9% accuracy, but it struggled with class imbalance. Satya Islam et al. [7] developed a hybrid prediction model combining K-Nearest Neighbor and Bernoulli Naive Bayes, achieving 92.86% classification accuracy, though the ensemble approach led to overfitting and complexity. Khan et al. [8] leveraged a Kaggle dataset to create an early disease detection model with 96% accuracy, but it lacked data preprocessing techniques. Alanazi et al. [9] introduced a CNN-KNN-based disease prediction system that also reached 96% accuracy but was unable to fully resolve class imbalance issues.

Given these limitations, more effective strategies are needed to improve feature selection, address class imbalance, and enhance model generalization for accurate and early CKD prediction.

3. PROBLEM STATEMENT

Accurately diagnosing chronic diseases using clinical decision support systems remains a complex task, as analyzing clinical and genetic datasets significantly affects model performance and efficiency. A single predictive model often faces multiple challenges in detecting renal disorders, including high dimensionality, suboptimal feature selection, imbalanced data distribution, prolonged training time, and increased computational costs.

The presence of missing values and outliers can negatively impact the model's overall effectiveness, while irrelevant or redundant features contribute to increased dimensionality and reduced predictive accuracy. Additionally, as processing time increases, the complexity of the classifier leads to higher resource consumption. These limitations collectively hinder the model's functionality and decrease its predictive capability.

To overcome these challenges, this research aims to develop a robust and efficient model for the early diagnosis of chronic kidney disease, improving classification accuracy, feature selection, and computational efficiency.

4. RESEARCH OBJECTIVES

* To develop a hybrid classifier capable of selecting the most relevant features from the dataset for precise classification.
* To enhance disease prediction accuracy by addressing high-dimensionality challenges and optimizing the feature learning process while minimizing computational overhead.
* To improve classifier performance by introducing a refined model that reduces training time, strengthens prediction reliability, and effectively recognizes complex patterns.

5. MOTIVATION OF THE RESEARCH

This paper aims to enhance disease prediction accuracy and efficiency to improve patient care while reducing healthcare costs. By developing models that enable faster and more precise diagnoses, the study seeks to support better clinical decision-making. The motivation behind this research lies in its potential to transform healthcare by providing advanced tools for early disease detection, ultimately leading to improved patient outcomes. The proposed models focus on optimizing feature learning to extract meaningful insights from data, resulting in a reliable and effective prediction and classification system.

6. RESEARCH CONTRIBUTIONS

This research is broken down into three levels, each of which focuses on a different component of model building and optimization for the early identification and precise categorization of chronic kidney disease. The block diagram for the contributions is displayed in

Figure 1.

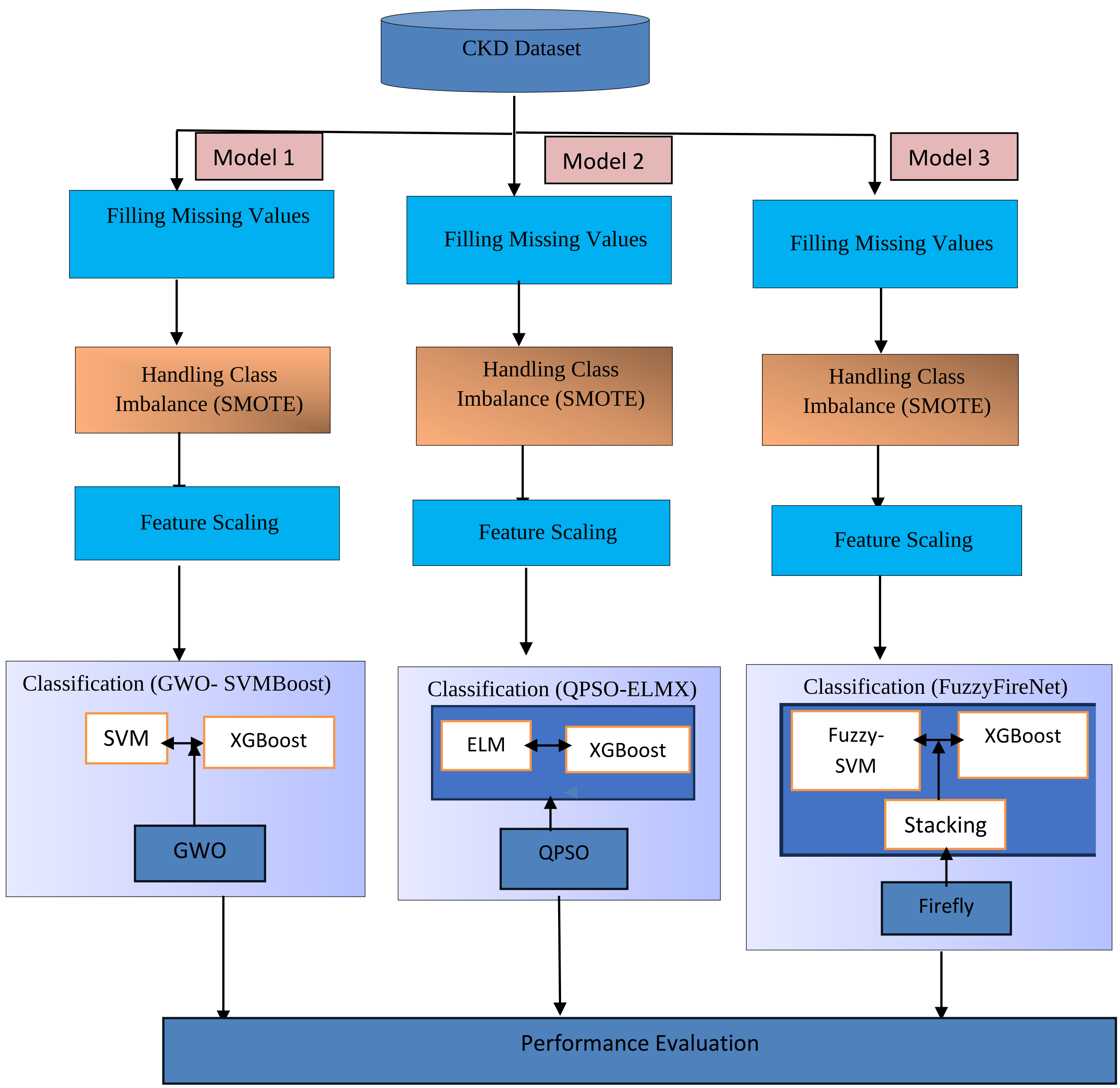

Figure 1. Block Diagram of the Proposed Model

Proposed Model 1: GWO-SVMBoost Prediction model

To achieve optimal classification, the proposed model integrates the Support Vector Machine (SVM) classifier and XGBoost with the Grey Wolf Optimization (GWO) feature selection technique for Chronic Kidney Disease (CKD) prediction. Preprocessing techniques, including data cleaning and handling missing values, are applied to enhance data quality and maintain dataset integrity. This process improves model accuracy while ensuring reliable predictions.

Feature extraction and selection are performed before feeding data into the hybrid model to improve classification accuracy. The GWO-based feature selection method enhances efficiency by eliminating redundant attributes. By combining SVM and XGBoost, the hybrid classifier effectively captures both linear and non-linear relationships, improving generalization and handling complex patterns.

To address class imbalance, the Synthetic Minority Oversampling Technique (SMOTE), also known as the Adaptive Weighting Method, is applied. In the initial preprocessing stage, data normalization is carried out using Min-Max Scaling, while missing values are imputed using the K-Nearest Neighbors (KNN) method. The second stage involves feature selection using GWO to identify the most relevant attributes. Finally, the third stage employs the hybrid SVM-XGBoost classifier for robust prediction.

Despite its advantages, the proposed model has limitations, including extended training times, scalability challenges, and high computational costs.

Proposed Model 2: Quantum Particle Swarm Optimized Extreme Learning Machine with XGBoost (QPSO-ELMX) Classifier

The proposed model enhances feature quality, manages high-dimensional data, and minimizes computational complexity by grouping similar properties of chronic kidney disease. To address class imbalance, the model utilizes SMOTE for data preprocessing. Additionally, Quantum Particle Swarm Optimization is implemented to identify and select the most relevant features, ranking them based on their correlation with the target variable.

For classification, the Extreme Learning Machine combined with XGBoost is employed, ensuring fast processing and improved accuracy after feature selection. This approach effectively mitigates issues related to high dimensionality and class imbalance, ultimately enhancing predictive performance in CKD detection. However, limitations of the model include challenges related to uncertainty, robustness, and the need for further optimization in feature selection techniques.

Proposed Model 3: FuzzyFireNet Classifier

The FuzzyFireNet Classifier model is proposed as a solution to the issues with the earlier models. The benefits of the Firefly Optimized Ensemble and Fuzzy Logic-SVM approaches for diagnosing chronic kidney disease are used by this proposed methodology. The three steps of this suggested approach are data pre-processing to eliminate outliers, KNN imputation for missing values, and Z-score normalization for standardizing the values. The SMOTE method is used to address the class imbalance. A certain target variable is used to rank each feature, and when a strong absolute correlation value is found, the features are retrieved. For improved stability and feature selection, the Firefly algorithm is used to choose the extracted features. For the optimal classification procedure, the FuzzyFireNet Classifier is proposed. In order to use machine learning techniques for improved classification problems, it combined the deep architecture with the classifiers Stacking Ensemble Model, Fuzzy Support Vector Machine, and XGBoost. The XGBoost model is trained for complicated pattern recognition, while fuzzy SVM is trained for handling uncertainty. The stacking ensemble learning method is used to integrate their predictions. In order to improve interpretability, accuracy, and training time, the FuzzyFireNet Classifier Model is a hybrid technique that combines the Firefly Optimised Ensemble approach with the Fuzzy Logic-Support Vector Machine. When compared to the other two suggested methods, the FuzzyFireNet Classifier Model methodology yields a more robust and accurate model for identifying chronic kidney disease with less processing time and a powerful and efficient mechanism for fine-tuning.

7. EXPERIMENTAL RESULTS

The proposed research on Chronic Kidney Disease (CKD) detection is implemented using Python within the PyCharm development environment. The system configuration consists of an Intel Core i5 processor, Windows 10 operating system, 8GB RAM, and a 512GB SSD. This study utilizes the Chronic Kidney Disease dataset available on Kaggle to facilitate early diagnosis. The dataset comprises 309 patient records, encompassing both diseased and non-diseased cases. Of these, 270 records (90%) correspond to non-diseased individuals, while 39 records (10%) represent patients diagnosed with CKD. Each record contains multiple attributes, represented as either numerical or categorical data. For analysis and prediction, 16 key features are selected.

The evaluation of the proposed methods is conducted based on accuracy and processing time. Table 1 illustrates the classification performance of three models: Grey Wolf Optimized SVMBoost (GWO-SVMBoost), Quantum Particle Swarm Optimized Extreme Learning Machine with XGBoost (QPSO-ELMX), and FuzzyFireNet. Among these, the FuzzyFireNet model exhibited the highest accuracy while maintaining the shortest training time compared to the other two approaches.

Table.1. Comparison Results for Proposed Models

Models	Kidney Disease Dataset			
	Accuracy (%)	Precision (%)	Recall (%)	Processing time (s)
Grey Wolf Optimized SVMBoost (GWO-SVMBoost)	**91.18**	**91.13**	**88.16**	**141.11**
Quantum Particle Swarm Optimized Extreme Learning Machine with XGBoost (QPSO-ELMX)	**92.26**	**98.64**	**98.61**	**140.09**
FuzzyFireNet	**96.52**	**97.14**	**96.68**	**137.12**

Table 1 presents the accuracy, precision, recall, and processing time for the proposed models. Among them, the FuzzyFireNet classifier achieved the highest accuracy of 96.52% with a processing time of 137.12 seconds on the chronic kidney disease dataset. Additionally, the Quantum Particle Swarm Optimized Extreme Learning Machine with XGBoost (QPSO-ELMX) and Grey Wolf Optimized SVMBoost

(GWO-SVMBoost) demonstrated superior performance compared to existing methods, further enhancing classification accuracy and efficiency.

8. CONCLUSION

This research paper tackles key challenges in the accurate detection and prediction of kidney diseases using advanced machine learning techniques. Three innovative models—Grey Wolf Optimized SVMBoost (GWO-SVMBoost), Quantum Particle Swarm Optimized Extreme Learning Machine with XGBoost (QPSO-ELMX), and FuzzyFireNet—were developed to improve predictive accuracy and efficiency. These models effectively addressed classification challenges, mitigated class imbalance, managed high-dimensional data, and overcame issues related to shallow feature learning and overfitting while optimizing processing time. By refining classification performance, the proposed models enhanced prediction accuracy, creating a more reliable framework for disease detection. These advancements contribute to medical diagnostics, support early clinical decision-making, and improve patient care outcomes.

9. REFERENCES

[1] Salman, R., & Gupta, S. (2023). Hybrid machine learning model for chronic disease prediction. International Journal of Intelligent Systems and Applications in Engineering (IJISAE), 11(2), 808–816.

[2] Lee, C., Jo, B., Woo, H., et al. (2022). Chronic disease prediction using the common data model: Development study. JMIR AI, 1(1), e41030.

[3] Othman, K. T. (2025). Modeling chronic disease prediction using statistical machine learning methods. Advances in Nonlinear Variational Inequalities, 28(5s).

[4] Kulkarni, K., Manjunath, B., Hebbar, M. T. M., et.al. (2021). Chronic disease prediction using machine learning. International Journal of Advanced Research in Computer and Communication Engineering, 10(6).

[5] Korke, V. K., Chowdhary, V., Keri, S. M., et al. (2022). Chronic disease prediction using machine learning. International Journal for Research in Applied Science & Engineering Technology (IJRASET), 10(VIII).

[6] Ramani, K., Suneetha, I., Pushpalatha, N., et al. (2022). Chronic disease prediction through supervised learning techniques. Journal of Pharmaceutical Negative Results, 13(4).

[7] Islam, R., Sultana, A., & Islam, M. R. (2024). A comprehensive review for chronic disease prediction using machine learning algorithms. Journal of Electrical Systems and Information Technology.

[8] Khan, A. (2022). Machine learning for chronic disease prediction. CEOS Public Health Research, 1(1), 101.

[9] Alanazi, R. (2022). Identification and prediction of chronic diseases using a machine learning approach. Journal of Healthcare Engineering, 2022, Article ID 2826127.