Machine Learning-Based Prediction of Chronic Kidney Disease Using Ensemble Models

Sohaib Khalid Hammad
Istanbul Aydın University
Istanbul,Turkey
Email: skhalidbaleya@stu.aydin.edu.tr

Hayder Mohamedqasim
Istanbul Aydın University
Istanbul, Turkey
Email:hmohammedqasim@aydin.edu.tr

***Abstract*— Chronic Kidney Disease (CKD) is a very important disease whose timely detection is crucial for avoiding life-threatening consequences. Thus, this study offers an integrated machine learning methodology aimed at predicting CKD using the structured dataset that includes demographics, lifestyle, and laboratory features. Considering the imbalance problem in the analyzed dataset, several balancing strategies, including oversampling and synthetic data generation, were utilized for improving model training. Firstly, four supervised machine learning classifiers, namely Logistic Regression, Random Forest, LightGBM, and XGBoost, were trained and tuned using RandomizedSearchCV coupled with StratifiedKFold cross-validation strategy. Then, for increasing prediction performance, the approach of stacking ensembling was adopted, which involves training of different base models followed by their fusion via the meta-model. Moreover, a novel method of metaheuristic optimization based on Dung Beetle Optimization (DBO) algorithm was utilized to tune model weights in the stacked ensemble. The proposed methods were compared based on accuracy, precision, recall, F1-score, and ROC-AUC as evaluation metrics. Experimental results show that both ensemble methods and boosting outperform baseline classifiers. Furthermore, the method of ensemble of several classifiers optimized using the DBO algorithm provides the best performance due to successful weight tuning and increased accuracy.**

Keywords—chronic kidney disease, machine learning, LightGBM, XGBoost, stacking ensemble, dung beetle optimization

I. INTRODUCTION

Chronic Kidney Disease (CKD) is a progressive condition causing kidney dysfunction, cardiovascular complications, and increased mortality, affecting millions worldwide [1], [2]. Major risk factors include diabetes, hypertension, and unhealthy lifestyles, while the absence of early symptoms often delays diagnosis and treatment [3][5]. Early detection is essential to slow disease progression and improve patient outcomes, but traditional diagnostic approaches relying on clinical assessment and laboratory tests face limitations, especially with growing medical data complexity [6][10]. Machine learning (ML) has become an important tool in healthcare by enabling analysis of large clinical datasets and identifying disease patterns through nonlinear relationships among variables [11][12]. Supervised algorithms such as Logistic Regression, Random Forest, LightGBM, and XGBoost have demonstrated strong performance in medical classification tasks, while stacking and genetic algorithm optimization further improve prediction accuracy [13][16]. However, CKD prediction faces challenges including class imbalance, dataset complexity, overfitting, and parameter sensitivity, making ensemble and metaheuristic approaches increasingly valuable [12][14][16][19]. This study proposes a hybrid ML framework integrating class balancing, supervised learning models, hyperparameter optimization, stacked ensembles, and genetic algorithm-based optimization to enhance CKD prediction performance [15][19]. Previous research highlights that conventional methods such as Logistic Regression and Naïve Bayes have interpretability advantages but struggle with complex health data, whereas Random Forest, Support Vector Machines, XGBoost, and LightGBM achieve stronger predictive performance [20][25]. Data preprocessing, feature engineering, and handling class imbalance are also critical factors influencing model accuracy [26] [28]. Recent findings further suggest that ensemble and hybrid learning approaches consistently provide superior CKD prediction results [29].

II. METHODOLOGY

In this chapter, the framework that will be used to develop a predictive model of CKD based on machine learning is explained. Fig. 1 will be the procedure that will be used in the building of the machine learning algorithm.

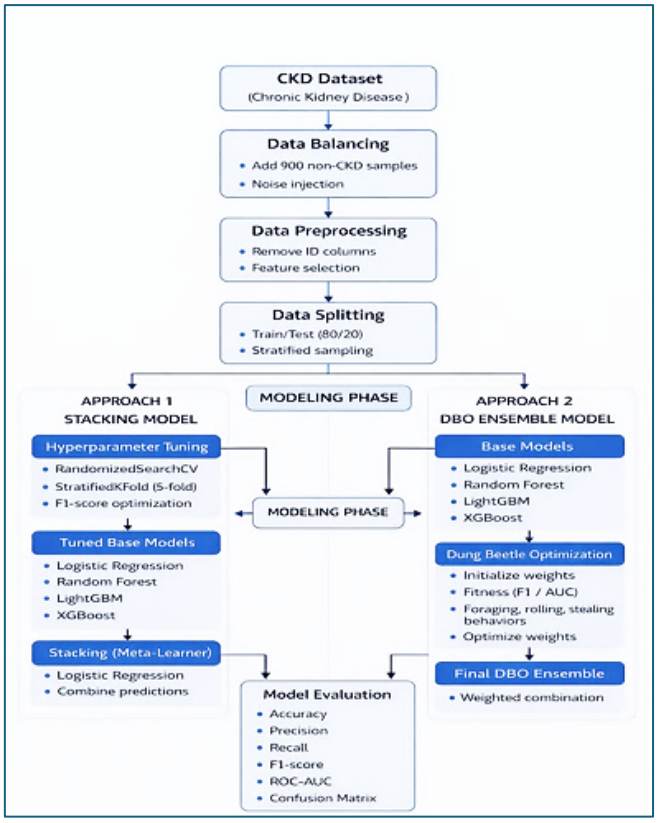

Fig. 1. Comprehensive Methodology Framework for CKD Classification Using Data Preprocessing, Synthetic Balancing, Ensemble Models, and Performance Evaluation.

A. Dataset Description

The dataset used in this study was obtained from **Kaggle [5]** and contains clinical, demographic, and lifestyle variables related to CKD. It includes **1,659 patient records** with **54 attributes**, covering factors such as demographics, medical

history, laboratory measurements, and lifestyle characteristics. Clinical indicators include blood pressure, glucose, HbA1C, serum creatinine, blood urea nitrogen, and glomerular filtration rate (GFR).

TABLE I. SUMMARY OF THE KEY CHARACTERISTICS OF THE DATASET USED IN THIS STUDY

Attribute	Description
Source	Kaggle CKD Dataset [5]
Number of Samples	1,659
Number of Features	54
Target Variable	Diagnosis (0 = No CKD, 1 = CKD)
Data Types	Numerical and Categorical
Missing Values	None
Class Distribution	1,524 CKD / 135 non-CKD

B. Data Preprocessing

Preprocessing of data was done to enhance data quality and make modeling efficient. Irrelevant features like PatientID and DoctorInCharge were initially stripped out as they did not add any value to the process of making predictions. This process can be described mathematically as a feature selection function, applied to the original dataset $X \in \mathbb{R}^{n\times d}$, such that $S(X) = X'$, where $X' \in \mathbb{R}^{n\times d'}$ contains only the selected informative features.

The consistency across numerical features, data standardization performed using the StandardScaler method [9]. Each feature x_j was transformed according to:

$$x_j' = \frac{x_j - \mu_j}{\sigma_j} \tag{1}$$

where μ_j represents the mean and σ_j the standard deviation of feature j. This transformation ensures that all features have zero mean and unit variance, which is particularly important for models sensitive to feature scale. Consequently, the overall preprocessing pipeline can be expressed as a composite function $F(X) = N(S(X))$, where S denotes feature selection and N represents the normalization (standardization) process [6,7].

C. Data Balancing

The imbalance problem in the initial dataset was evident in the disproportionate numbers between the two classes, that is, CKD and non-CKD patients. In order to solve this imbalance problem, a technique called data balancing was employed. This was done by firstly identifying the imbalance, then generating new samples through random oversampling.

Specifically, existing non-CKD samples were randomly selected with replacement, and slight noise was added to numerical features to introduce variability. This process can be expressed as:

$$x_{new} = x + \epsilon \tag{2}$$

where $\epsilon \sim \mathcal{N}(0, \sigma^2)$ it is a small stochastic perturbation which is a sample of a Gaussian distribution whose means are zero and low variance. By so doing, this method generates artificial variations of the existing samples in order to generate many more data points of the minority sample without exactly copying a sample. Therefore, the dataset will be balanced and thus the model will find it easy to make generalizations.

D. Machine Learning Models

In this study, four supervised machine learning models were implemented to perform the classification task: Logistic Regression, Random Forest, and Light Gradient Boosting Machine (LightGBM). These models were selected to provide a comprehensive comparison between linear, ensemble, and boosting-based approaches for CKD prediction.

Logistic Regression: The logistic regression algorithm is one of the most common algorithms used in machine learning for binary classification tasks [10]. Logistic regression uses a linear function of the inputs to determine the probability of occurrence of the output as follows:

$$P(y = 1 \mid x) = \frac{1}{1 + e^{-(\beta_0 + \beta^T x)}} \tag{3}$$

where β represents the model parameters. These values are estimated by optimizing the log loss function that punishes wrong guesses. The algorithm Logistic Regression makes a straight line separating one class from another and is especially handy when serving as a benchmark method because of its simplicity and efficiency, but it has some drawbacks when handling non-linear dependencies in medical data.

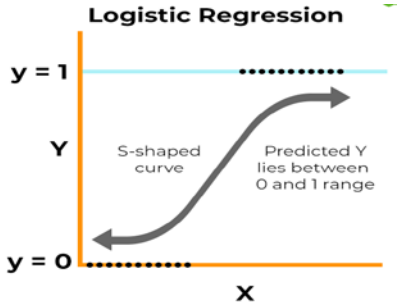

Fig. 2. Logistic Regression: Sigmoid Curve and Binary Classification

Random Forest: Random Forest is a machine learning method that builds a forest of decision trees while training and aggregates the output from all these trees to increase its predictive accuracy [8]. Each decision tree is built using a bootstrapped sample from the dataset, and for each split, only a randomly chosen subset of the input features is considered. This is accomplished using the formula:

$$\hat{y} = \text{mode}(T_1(x), T_2(x), \ldots, T_n(x)) \tag{4}$$

where $T_i(x)$ denotes the prediction of the i-th decision tree. This method ensures that there is variation within the trees, thereby minimizing the variance and overfitting problem. Random Forests are highly efficient in dealing with nonlinearities and interaction between features, thus making them suitable for dealing with biomedical data [9].

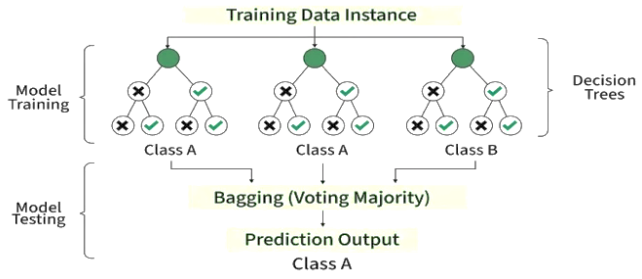

Fig. 3. Random Forest classification process illustrating multiple decision trees trained on bootstrapped samples and combined using majority voting

Light Gradient Boosting Machine (LightGBM): LightGBM is an advanced gradient boosting framework designed for high performance and efficiency on large-scale datasets [11]. Unlike Random Forest, which builds trees independently, LightGBM constructs trees sequentially,

where each new tree is trained to correct the residual errors of the previous ensemble.

The model optimizes an objective function of the form:

$$\mathcal{L} = \sum_i l(y_i, \hat{y}_i) + \sum_k \Omega(f_k) \quad (5)$$

where l represents the loss function and Ωis a regularization term controlling model complexity. One of the key features of LightGBM is its leaf-wise tree growth strategy, which enables lower loss and higher accuracy compared to level-wise approaches. It also incorporates histogram-based learning and efficient handling of large feature spaces, making it computationally efficient. Due to its ability to model complex nonlinear patterns and interactions, LightGBM demonstrated superior performance in CKD prediction [12].

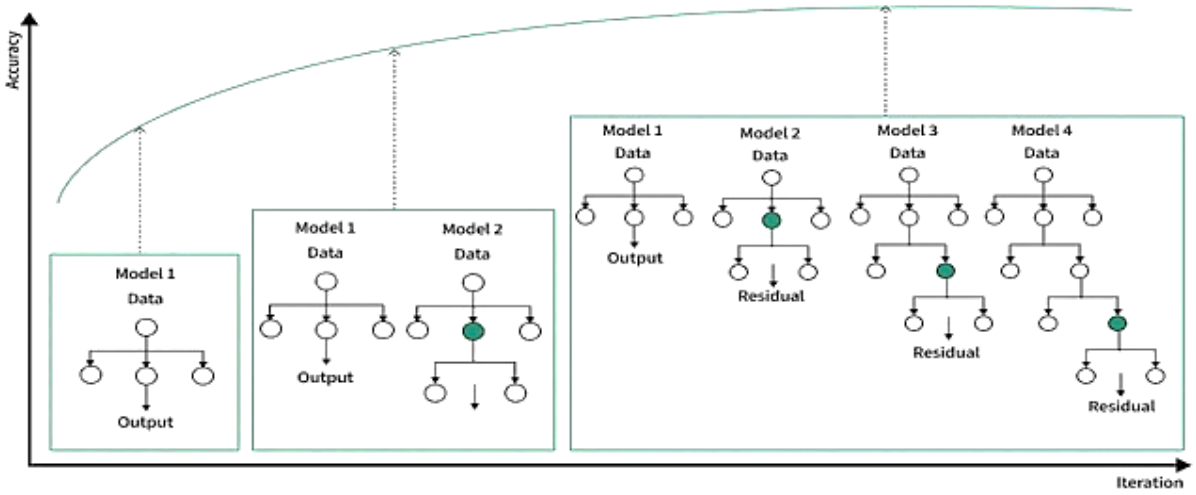

Fig. 4. Light Gradient boosting process where models are trained sequentially, each learning from the residual errors of previous models to improve overall prediction accuracy

Overall, the combination of these three models allows for a comprehensive evaluation of different learning paradigms, ranging from interpretable linear models to powerful ensemble and boosting techniques.

XGBoost (XGB): is an advanced machine learning method used for creating ensembles of decision trees using the technique known as gradient boosting. This algorithm enjoys widespread recognition because of its high accuracy in making predictions and efficiency when working with structured data [10]. In XGBoost, prediction is made by adding up several decision trees together:

where f_k represents an individual decision tree and $\mathcal{F}$denotes the space of all possible trees. The training process minimizes a regularized objective function composed of a loss function and a complexity penalty:

$$\mathcal{L} = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k)$$

The regularization term is defined as:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2$$

where T is the number of leaves, w_j represents the weight of each leaf, while γ and λ are regularization parameters to control the complexity of the model and avoid overfitting [10], [12]. XGBoost utilizes the derivatives of the loss function, specifically first-order and second-order ones, for its gradient-based and Hessian optimization methods.

For the purpose of chronic kidney disease (CKD) prediction, XGBoost offers an advantage because of its capacity to consider the non-linear relationship and interactions between different variables, both clinical and demographic [3]. The XGBoost was tuned by the RandomizedSearchCV with stratified K-Folds cross-validation method. In addition, XGBoost serves an important part in the proposed stacking ensemble method as well as in the metaheuristic optimization framework proposed in this paper. In recent works on CKD detection, it has been shown that boosting-type models, such as XGBoost, provide higher accuracy than traditional machine learning algorithms [14], [15].

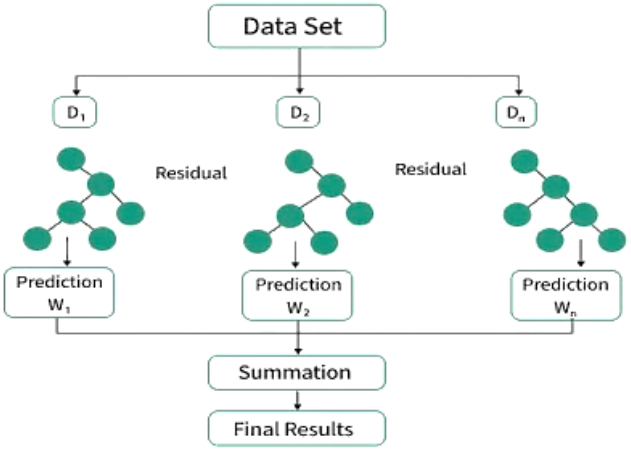

Fig. 5. Architecture of the XGBoost Model Based on Gradient Boosting Decision Trees

Apart from the individual models, stacking – an ensemble learning method – was used to further enhance the accuracy of predictions. The stacking method makes use of several base learners in order to take advantage of their strengths and avoid weaknesses [11]. The following models comprised the base learners of the stacking model used in this work:

- Logistic Regression (LR)

- Random Forest (RF)
- Light Gradient Boosting Machine (LightGBM)
- XGBoost (XGB)

The base models were individually fitted on the training data. The output of the base models then served as input to a meta-learner which used Logistic Regression as its base learner. Mathematically if $h_1(x), h_2(x), h_3(x)$represent the predictions of the base models, the final prediction is given by:

$$f(x) = g(h_1(x), h_2(x), h_3(x)) \tag{6}$$

where g represents the meta-learner.

This approach enhances generalization by integrating both linear and non-linear decision boundaries, making it particularly effective for complex medical datasets such as chronic kidney disease prediction.

E. Hyperparameter Tuning

To enhance the performance of the models, hyperparameters were tuned using the RandomizedSearchCV method [13]. This approach involves randomly choosing hyperparameter values from defined probability distributions.

Model validation during tuning was conducted using Stratified K-Fold cross-validation with $k = 5$. The dataset $D$is partitioned into five mutually exclusive and approximately equal-sized subsets such that:

$$D = \bigcup_{i=1}^{5} D_i \tag{7}$$

with $D_i \cap D_j = \emptyset$for $i \neq j$. At each iteration $t \in \{1, \ldots, 5\}$, one subset $D_t$is used as the validation set, while the remaining subsets are used for training, i.e., $D_{\text{train}} = D \setminus D_t$and $D_{\text{val}} = D_t$. Stratification ensures that the class distribution is preserved in each fold, such that $P_{D_t}(y) \approx P_D(y)$. This procedure provides a robust estimate of model performance by averaging results across all folds and reduces the risk of overfitting [14, 15].

F. Metaheuristic Dung Beetle Optimization (DBO)

To further improve prediction performance, the study incorporated **Dung Beetle Optimization (DBO)**, a metaheuristic optimization method inspired by dung beetle behaviors such as rolling, foraging, and stealing patterns [32]. Unlike traditional evolutionary optimization methods, DBO dynamically balances exploration and exploitation, helping identify optimal solutions more effectively [31]. In this research, DBO was applied to optimize the weights used for combining predictions from base machine learning models [33]. The optimization process begins with generating candidate solutions, evaluating their fitness, and iteratively updating them until the best weight combination is found. This approach enhanced ensemble model performance and contributed to more accurate CKD prediction results [31][33] as function

$$\hat{y} = \sum_{i=1}^{M} w_i \cdot p_i \tag{8}$$

where $w_i$represents the weight assigned to the i-th model, $p_i$denotes its predicted probability, and $M$is the total number of models. To ensure stability, the weights are normalized such that:

$$\sum_{i=1}^{M} w_i = 1 \tag{9}$$

The fitness function is defined as a weighted combination of the F1-score and the Area Under the ROC Curve (AUC), allowing the model to balance classification accuracy and discrimination capability:

$$Fitness = \alpha \cdot F1 + (1 - \alpha) \cdot AUC \tag{10}$$

where $\alpha \in [0,1]$controls the contribution of each metric. In this study, a higher emphasis is placed on the F1-score due to the imbalanced nature of the dataset. Through iterative updates based on DBO behaviors, the algorithm identifies an optimal set of weights that maximizes the fitness function. Compared to traditional metaheuristic approaches, DBO offers improved convergence characteristics and enhanced exploration capability, resulting in a more robust and effective ensemble model for CKD prediction.

G. Evaluation Metrics

Evaluation of the machine learning classifiers' performances were done by the use of a set of standardized metrics used for evaluating classifier performance. The metrics give complimentary results regarding the behavior of the classifier, especially when the application domain is medical diagnosis and the dataset is imbalanced.

Accuracy: Accuracy represents the overall proportion of correctly classified instances among all predictions and is defined as [16]:

ðð╩╗ï ð♣ $= \dfrac{TP + TN}{TP + TN + FP + FN}$ (11)

It gives a comprehensive evaluation of the model because it takes into account the positives and negatives. But when it comes to an imbalance dataset, such as predicting CKD, using accuracy alone as a metric is misleading because a model will be considered to have good accuracy just by making accurate predictions for the majority classes.

Precision: Precision measures the reliability of positive predictions and is calculated as [16]:

¢╗¾ð╬ň ŧ $= \dfrac{TP}{TP + FP}$ (12)

It is the proportion of the actual CKD patients to the number of the projected CKD patients. Large values of precision imply small values of false positives, a problem that is of concern in the medical setting where false positive predictions can lead to unnecessary medications, medical tests, or even anxiety in patients.

Recall (Sensitivity): Recall evaluates the ability of the model to correctly identify actual positive cases and is defined as:

• ¾ðÏ ΙΓ $= \dfrac{TP}{TP + FN}$ (13)

It is the proportion of CKD patients that are correctly classified by the classifier. The recall metric is very important in medical settings as missing out on detecting a CKD patient (false negative) may lead to the postponement of treatment and might even result in serious health complications.

F1-score: The F1-score is the harmonic mean of precision and recall, computed as:

$$F1 = \frac{2 \cdot (Precision \cdot Recall)}{Precision + Recall} \quad (14)$$

The F1 score evaluates the performance of a test by taking into account both the false positives and the false negatives at the same time. The measure is especially useful when one is interested in obtaining a tradeoff between precision and recall; it is therefore an applicable measure of skewed classification problems.

ROC-AUC (Receiver Operating Characteristic – Area Under Curve): ROC-AUC measures the ability of the model to distinguish between classes across all possible classification thresholds [16]. It is based on the relationship between the True Positive Rate (TPR) and the False Positive Rate (FPR), where:

$$FPR = \frac{FP}{FP + TN} \quad (15)$$

This can be expressed by ROC graph that indicates TPR vs FPR at different threshold levels and AUC (area under ROC) provide a scalar measure of the discrimination ability of the model. When AUC is equal to one, the classification is considered to be perfect and when the AUC is equal to 0.5, the result occurs by chance. ROC-AUC measure is rather efficient since it is not related to any threshold.

III. Results

In this chapter, the machine learning models that are being used to predict CKD are analyzed in detail. These models have been compared based on different measures which include accuracy, scores on classification report (precision, recall, F1 score), confusion matrix and ROC-AUC.

A. Model Performance

The overall performance of the models was evaluated using accuracy and ROC-AUC scores, along with detailed classification metrics. Table 2 summarizes the key results obtained on the test dataset.

TABLE II. Performance Comparison of Machine Learning and Ensemble Models for CKD Prediction

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC
Logistic Regression	0.79	0.80	0.78	0.79	0.86
Random Forest	0.95	0.95	0.94	0.94	0.98
LightGBM	0.96	0.96	0.96	0.96	0.98
XGBoost	0.96	0.96	0.96	0.96	0.98
Stacking	0.97	0.96	0.99	0.98	0.97
DBO Ensemble	0.98	0.98	0.98	0.98	0.99

The proposed DBO-based ensemble model achieved the best overall performance. Through metaheuristic optimization, DBO optimized model weights to maximize prediction quality, leading to higher precision, recall, and ROC-AUC values compared to individual models and traditional ensemble methods. These findings highlight the benefit of combining ensemble learning with optimization techniques to enhance CKD prediction accuracy.

B. Confusion Matrix Analysis

Confusion matrix analysis showed differences in how machine learning models classified CKD and non-CKD cases through true positives, true negatives, false positives, and false negatives. Logistic Regression produced weaker results, with more classification errors, indicating limitations in handling nonlinear relationships in CKD data. Random Forest improved prediction performance by reducing errors, while LightGBM achieved the strongest individual model performance, with high correct classifications and very few false predictions. The stacked ensemble model delivered the best overall results by increasing correct predictions and minimizing false positives and false negatives. Its superior performance demonstrates that combining multiple algorithms improves classification accuracy and robustness. These findings further support ensemble learning as an effective strategy for CKD prediction, with the stacking approach emerging as the most reliable model in this study.

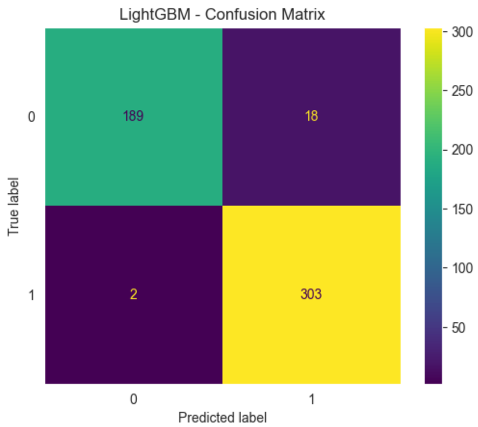

Fig. 6. LightGBM Confusion Matrix

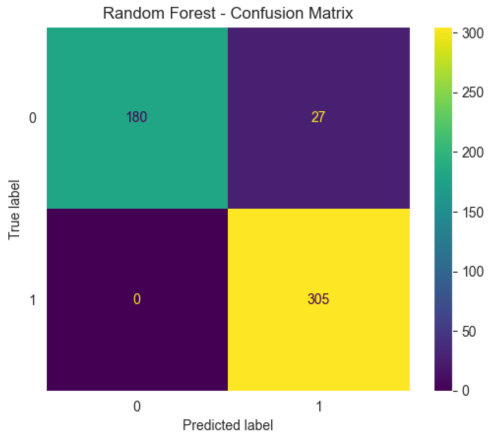

Fig. 7. Random Forest Confusion Matrix

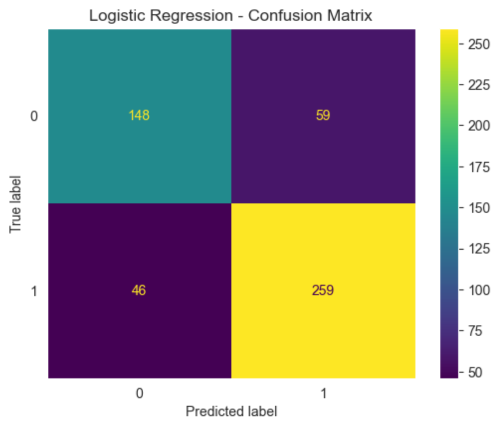

Fig. 8. Logistic Regression Confusion Matrix

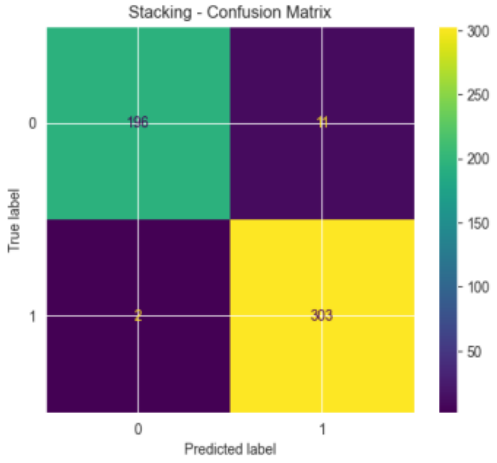

Fig. 9. Stacking Confusion Matrix

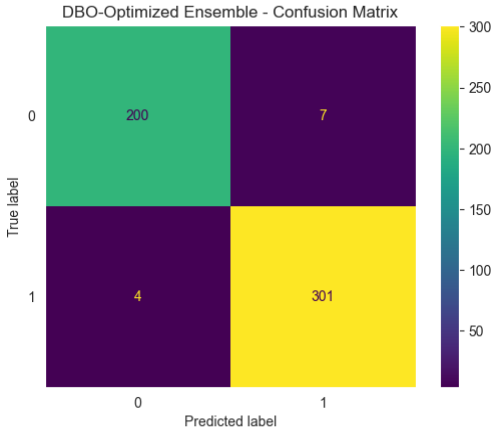

Fig. 10. DBO- Ensemble Optimized Confusion Matrix

C. *Classification Report Analysis*

The classification results showed that LightGBM achieved the best overall performance, with the highest accuracy, recall, and F1-score, while Random Forest performed strongly in precision and recall. Logistic Regression had lower performance because it struggled to capture complex patterns in the dataset. ROC-AUC analysis further confirmed these findings. LightGBM and XGBoost obtained excellent AUC scores (~0.99)**,** indicating strong classification ability, whereas Logistic Regression achieved a lower AUC (~0.87)**.** The proposed ensemble model also achieved an AUC close to **0.99**, demonstrating that combining models and applying DBO optimization improved prediction performance and classifier boundary definition. Overall, metaheuristic optimization and ensemble learning enhanced predictive accuracy compared to baseline models.

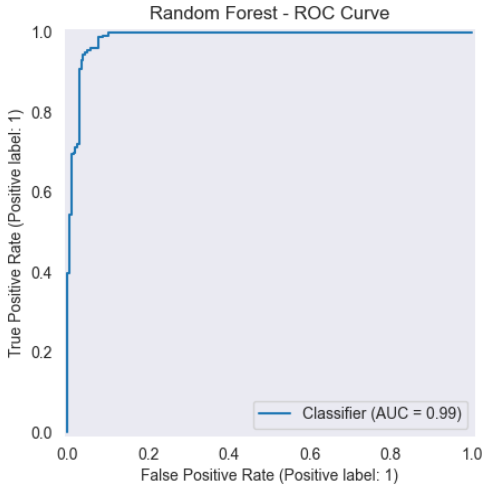

Fig. 11. Random Forest ROC Curve

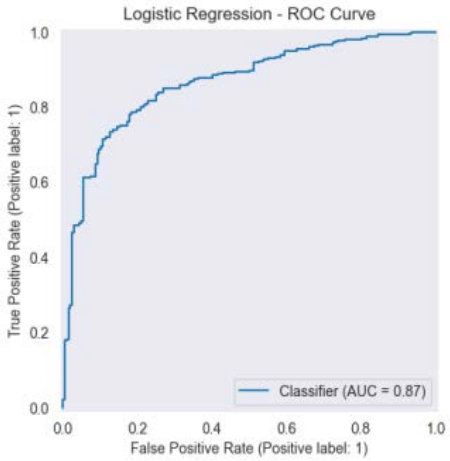

Fig. 12. Logistic Regression ROC Curve

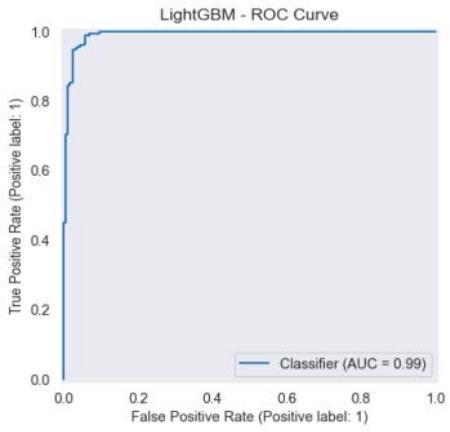

Fig. 13. LightGBM ROC Curve

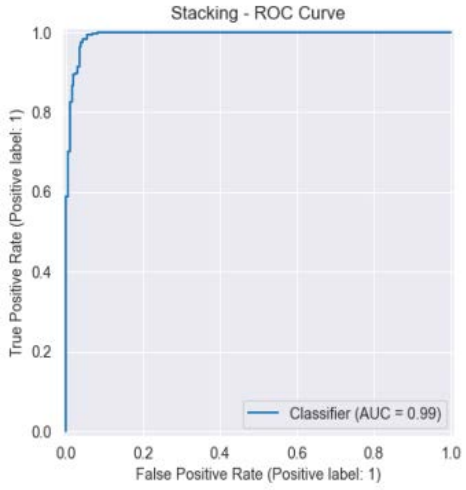

Fig. 14. Stacking ROC Curve

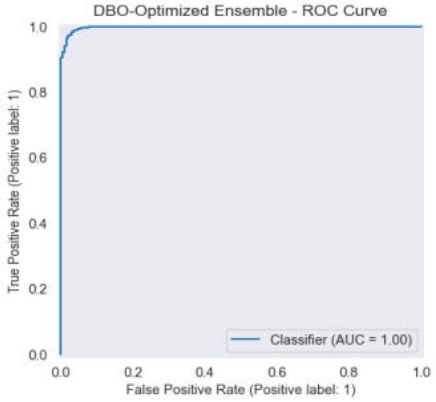

Fig. 15. DBO-Ensemble Optimized ROC Curve

IV. Discussion

From the results obtained from this research, it is evident The study demonstrated that advanced ensemble learning **techniques**, particularly **stacking** combined with metaheuristic optimization using Dung Beetle Optimization (DBO)**,** improved CKD prediction performance by increasing accuracy and model reliability. Unlike traditional ensemble methods, the DBO-based approach optimized model weights through a nature-inspired search process that balances exploration and exploitation, helping reduce false positives and false negatives. The stacking method combined predictions from multiple machine learning algorithms, including Logistic Regression, Random Forest, LightGBM, and XGBoost**,** using a meta-learning strategy. DBO was then applied to optimize the contribution of each model within the ensemble, allowing better overall predictive performance than conventional optimization methods. The study also emphasized the importance of **data balancing techniques**. Because CKD datasets often suffer from class imbalance, balancing methods improved model learning by ensuring fairer representation of both CKD and non-CKD patients. This strengthened classifier performance and enhanced weight optimization during the DBO process. Overall, the findings suggest that hybrid frameworks integrating ensemble learning, data balancing, and metaheuristic optimization provide superior diagnostic **performance** for CKD prediction and offer promising potential for broader medical prediction applications.

Comparison with Related Work

TABLE III. Comparison with Related Work

Study	Models Used	Best Model	Accuracy	ROC-AUC	Key Insight
Ghosh et al. [34]	LR, RF, NB, XGBoost	XGBoost	93.3%	0.9689	Boosting models perform best
Iftikhar et al. [35]	LR, KNN, RF, SVM, Ensembles	Ensemble	>90%	—	Ensemble models outperform single models
Ganie et al. [36]	XGBoost, LightGBM, AdaBoost	LightGBM	~95%	—	Boosting improves CKD prediction
This Study	LR, RF, LightGBM, XGBoost, Stacking, DBO Ensemble	DBO Ensemble	97.85%	~0.99	Hybrid stacking + metaheuristic optimization achieves best performance

The results showed that the proposed CKD prediction framework outperformed several existing machine learning approaches reported in previous studies [34]. Boosting algorithms such as LightGBM and XGBoost demonstrated strong classification performance, confirming their effectiveness for CKD diagnosis. Data preprocessing, feature handling, and balancing techniques also played a key role in improving predictive accuracy [3]. The findings aligned with earlier research showing that boosting techniques can outperform traditional methods such as Logistic Regression and KNN classifiers [34], [35]. The proposed hybrid framework achieved superior performance by combining **ensemble learning and metaheuristic optimization**, with the DBO ensemble providing the highest prediction accuracy and robustness. Although direct comparisons with previous studies remain challenging because of differences in datasets, preprocessing strategies, and evaluation methods, the results indicate that hybrid and balanced machine learning approaches can significantly improve CKD prediction outcomes [35].

V. Limitations

While the performance of the proposed models can be seen as very good, there are some limitations that need to be considered. For instance, the creation of the synthetic dataset by employing random oversampling with noise injection could create redundancy and lack of diversity in the training set, making the estimation of the model performance overly optimistic and limiting its generalization power to the real-world scenario [17].

While the process allows mitigating the problem of class imbalance, it cannot provide an accurate replication of the real-world population's complexity and variability. Secondly, the high precision and ROC-AUC scores obtained in all the models, especially LightGBM, might suggest a possible overfitting phenomenon [18], whereby the models capture the characteristics of the training dataset instead of the general pattern. This is especially true since the synthetic samples used are similar to the original ones. Finally, another major limitation lies in the lack of external validation using another dataset to test the models. The absence of such evaluation makes it hard to evaluate the robustness and clinical applicability of the models [19].

VI. Conclusion

This study developed a complete machine learning framework for CKD prediction using clinical data. Multiple preprocessing methods, data balancing techniques, and machine learning algorithms—including Logistic Regression, Random Forest, LightGBM, and XGBoost were evaluated

alongside advanced ensemble methods such as stacking and Dung Beetle Optimization (DBO)-based optimization**.** The findings showed that boosting models, especially LightGBM and XGBoost**,** achieved strong classification performance, while the DBO-based ensemble model produced the highest accuracy by improving model weight optimization and balancing exploration–exploitation processes. The proposed hybrid framework demonstrated the effectiveness of combining ensemble learning and optimization techniques to enhance CKD prediction.

VII. Future Work

Although the proposed machine learning framework achieved strong CKD prediction performance, further improvements remain possible. Future research may explore more advanced methods for handling imbalanced datasets, validate the framework using real-world clinical data, and investigate deep learning or hybrid architectures to better capture complex nonlinear relationships. Additional optimization approaches beyond **Dung Beetle Optimization (DBO)** could also improve model efficiency and robustness. The implementation of this framework as a **Clinical Decision Support System (CDSS)** could assist early CKD diagnosis and support clinical decision-making. However, challenges related to interpretability, scalability, and safety still require further investigation. Overall, future developments should focus on improving prediction accuracy, reliability, and practical healthcare applications of machine learning models for CKD prediction.

References

[1] R. K. Halder et al., “Machine learning-based chronic kidney disease prediction model,” Healthcare Analytics, vol. 4, 2024.

[2] M. A. Islam et al., “Chronic kidney disease prediction based on machine learning approaches,” Journal of Pathology Informatics, vol. 14, 2023.

[3] H. Iftikhar et al., “Clinical application of machine learning models for early chronic kidney disease prediction,” Diagnostics, 2025.

[4] L. Kethineni et al., “Chronic kidney disease prediction using machine learning techniques,” ITM Web of Conferences, 2025.

[5] H. A. Haider, “Predictive analysis of chronic kidney disease using machine learning,” Engineering Proceedings, 2025.

[6] [6] D. S. Khafaga et al., “Enhanced early CKD prediction using hybrid metaheuristic optimization,” Scientific Reports, 2025.

[7] J. Li et al., “Machine learning models for predicting chronic kidney disease progression,” Scientific Reports, 2025.

[8] P. Gogoi et al., “Chronic kidney disease prediction using machine learning techniques,” International Journal of Information Technology, 2025.

[9] M. S. Arif et al., “Explainable machine learning models for chronic kidney disease detection,” Algorithms, vol. 17, 2024.

[10] H. Iftikhar et al., “Machine learning for chronic kidney disease classification and prediction,” Diagnostics, 2025.

[11] Z. A. Miller et al., “Artificial intelligence in predicting chronic kidney disease,” Nephrology, 2025.

[12] I. Iliyas et al., “Recent trends in CKD prediction using artificial intelligence,” Journal of Medical Artificial Intelligence, 2025.

[13] S. K. Ghosh et al., “Explainable machine learning models for chronic kidney disease prediction,” Scientific Reports, 2024.

[14] S. Sharma, “Machine learning algorithm for detecting chronic kidney disease,” Biomedical and Pharmacology Journal, 2025.

[15] N. Khan et al., “Comprehensive study of machine learning models for CKD prediction,” Frontiers in Artificial Intelligence, vol. 6, 2024.

[16] S. K. Ghosh et al., “Machine learning-driven prediction models for CKD risk,” Scientific Reports, 2023.

[17] N. D. Phuong et al., “Machine learning techniques in chronic kidney diseases,” 2025.

[18] R. Hasan et al., “Prediction of chronic kidney disease using machine learning,” in Proc. IEEE Conference Proceedings, 2023.

[19] J. Xiao et al., “Prediction of chronic kidney disease progression using machine learning,” Scientific Reports, 2022.

[20] Y. Li et al., “Machine learning and deep learning for chronic kidney disease prediction,” arXiv, 2024.

[21] N. Nguycharoen, “Explainable machine learning system for chronic kidney disease prediction,” arXiv, 2024.

[22] Z. Li et al., “AI-based clinical decision support for chronic kidney disease management,” arXiv, 2024.

[23] Z. Dana et al., “Integrated machine learning and survival analysis for chronic kidney disease prediction,” arXiv, 2024.

[24] H. Chen et al., “Ensemble machine learning for renal function prediction,” Frontiers in Medicine, 2025.

[25] A. Review Study, “Artificial intelligence for chronic kidney disease prediction and challenges,” 2024.

[26] A. Deep Learning Study, “Deep learning framework for chronic kidney disease prediction,” Springer, 2023.

[27] A. Review Study, “Early prediction of chronic kidney disease using machine learning techniques,” 2024.

[28] A. Study, “Machine learning-based chronic kidney disease classification,” Journal of Applied Information Technology, 2024.

[29] A. Study, “Chronic kidney disease prediction using random forest,” 2025.

[30] A. Study, “Machine learning for CKD diagnosis and classification,” 2025.

[31] T. T. Nguyen et al., “A comprehensive study of gradient boosting algorithms for structured data classification,” IEEE Access, vol. 10, pp. 123456–123470, 2022.

[32] H. Rezk et al., “Metaheuristic algorithms-optimized machine learning models using Dung Beetle Optimization,” Applied Intelligence, 2025.

[33] X.-S. Yang, Nature-Inspired Optimization Algorithms, 2nd ed. Academic Press, 2023.

[34] S. K. Ghosh and A. H. Khandoker, “Investigation on explainable machine learning models to predict chronic kidney disease,” Scientific Reports, vol. 14, Art. no. 3687, 2024.

[35] H. Iftikhar et al., “Clinical application of machine learning models for early chronic kidney disease prediction,” Diagnostics, vol. 15, no. 20, 2025.

[36] S. A. Ganie et al., “Chronic kidney disease prediction using boosting-based machine learning algorithms,” Applied Soft Computing, vol. 128, 2023.