World Journal of Advanced Engineering Technology and Sciences
eISSN: 2582-8266
Cross Ref DOI: 10.30574/wjaets
Journal homepage: https://wjaets.com/

(RESEARCH ARTICLE)

Check for updates

Hybridized light gradient boosting and whale optimization algorithm for diabetes detection

Emmanuel Gbenga Dada [1, 2, *], Aishatu Ibrahim Birma [1], Abdulkarim Abbas Gora [1], Oluwasogo Adekunle Okunade [3] and Abubakar Hassan [4]

[1] *Department of Mathematics and Computer Science, Faculty of Science, Borno State University, Maiduguri.*
[2] *Department of Computer Science, Faculty of Physical Sciences, University of Maiduguri, Maiduguri, Nigeria.*
[3] *Department of Computer Science. Faculty of Computing, National Open University of Nigeria.*
[4] *Department of Computer Engineering, University of Maiduguri, Maiduguri, Borno State.*

World Journal of Advanced Engineering Technology and Sciences, 2025, 15(02), 1966-1981

Publication history: Received on 04 April 2025; revised on 13 May 2025; accepted on 15 May 2025

Article DOI: https://doi.org/10.30574/wjaets.2025.15.2.0760

Abstract

Due to their remarkable precision and effectiveness, gradient-boosted tree models have become the go-to choice for machine learning-driven diabetes detection; however, the key to unlocking their full potential lies significantly in the careful tuning of hyperparameters. To automatically optimize LGBM's hyperparameters for improved diabetes screening, we present a hybrid framework - Light Gradient Boosting (LGBM) bundled with the Whale Optimization Algorithm (LGBM+WOA). Inspired by nature, the Whale Optimization Algorithm (WOA) models the bubble-net feeding behaviour of humpback whales, therefore offering a compromise between exploration and exploitation in search areas. We evaluated model performance under imbalanced class situations using stratified 10-fold cross-validation using the Diabetes Dataset from patients in Borno hospital. Rising above baseline Gradient Boosting (80%), Support Vector Machine (74%), Random Forest (86%), and LGBM (88%), the suggested LGBM+WOA model achieved an overall detection accuracy of 90%. While diabetes recall increased to 0.86, so lowering false negatives is important; class-specific metrics for the non-diabetic cohort obtained a precision of 0.93, recall of 0.91, and F1-score of 0.92 - gains of 1-2 percentage points over standard LGBM. Faster convergence and better generalization follow from WOA-driven hyperparameter tuning, refining important LGBM parameters more effectively than grid or random search. The easier training and testing process of the hybrid model is a helpful tool for quickly assessing diabetes risk and allows for immediate use in clinical decision support systems. Combining LGBM's gradient-boosting efficiency with WOA's robust global optimization, the LGBM+WOA framework provides a new benchmark for machine-learning-based diabetes detection, enabling more general uses of metaheuristic-tuned ensembles in medical diagnostics.

Keywords: Diabetes; Support Vector Machine; Random forests; Light gradient boosting; Whale Optimization Algorithm

1. Introduction

Diabetes, formally Diabetes mellitus, is widely recognized as a chronic metabolic disorder primarily characterized by hyperglycemia, which occurs when a person has high blood sugar levels that can lead to complications such as blindness, cardiovascular diseases, and amputation [1]. The high blood glucose levels are due to insufficient insulin production, impaired insulin action or a combination of both factors. In the condition of diabetes, a patient's body is unable to generate sufficient insulin or to stop producing insulin [2]. Drugs alone, including insulin injections, are not sufficient to treat and cure diabetes. Scientists have not discovered the cure for the disease, but it can be controlled with early diagnosis and prognosis at the early stages of the disease and at the later stages of the disease which makes treatment

* Corresponding author: Emmanuel Gbenga Dada

much easier. The prediction of diabetes has become a controversial topic for study and research. The rapid progress of machine learning models has led to their widespread utilization in numerous applications, particularly in the medical field for the accurate diagnosis of diverse diseases [3]. To make diagnoses less costly and more accurate, there is a need for Machine learning. In general, machine learning models aim to describe, learn and predict from data and can help people make early judgments about disease based on their physical condition and diagnose the disease in its early stages until treatment is complete [3]. Therefore, using machine learning (ML) to predict diabetes can help doctors diagnose patients more efficiently and precisely [2].

Diabetes has emerged as one of the most public health challenges of the modern era, with its prevalence growing at an unprecedented rate. According to recent estimates, approximately 537 million adults were living with diabetes in 2021, with projections indicating that this number may rise to 643 million by 2030 and reach nearly 783 million by 2045 [4]. These statistics underscore the pervasive nature of the disease, especially in regions where rapid urbanization and lifestyle changes have contributed to a rise in risk factors such as obesity, sedentary behaviour, and unhealthy dietary habits. The economic implications of this rising epidemic are equally significant. Diabetes not only imposes a substantial financial implication on health care systems due to the costs associated with long-term management and treatment of its complications, but also exerts significant damage on national economies through lost productivity and premature mortality [5]. High-impact journals and peer-reviewed studies have consistently demonstrated that the financial strain imposed by diabetes affects both direct medical expenditures and indirect costs related to disability and reduced workforce participation. In recent years, there has been a notable shift toward leveraging advanced data analytics and machine learning to better understand and predict the epidemiological trends associated with diabetes. These studies show how machine learning can change our understanding of diabetes by revealing the complicated ways that genetics, environment, and behaviour interact. Overall, the global impact and epidemiology of diabetes reveal a multifaceted challenge that extends beyond individual health, affecting societies and economies worldwide [6]. The integration of machine learning techniques in epidemiological research not only enriches our understanding of the disease's distribution and progression but also informs the development of more effective, targeted strategies for early detection and intervention. As the global burden of diabetes continues to escalate, these innovative approaches are essential for mitigating the long-term consequences of the disease and optimizing resource allocation in public health initiatives. Early detection is paramount for reducing the incidence of severe complications associated with diabetes [7]. Traditional diagnostic methods such as fasting plasma glucose (FPG), oral glucose tolerance tests (OGTT), and hemoglobin A1c (HbA1c) assessments remain standard in clinical practice [8]. However, these techniques have inherent limitations. They typically offer a snapshot of glycemic control rather than capturing the dynamic, progressive nature of the disease. Also, they may not be very good at detecting early changes in metabolism, which could slow down the start of preventive actions [9].

Recent advances in machine learning have paved the way for more refined and anticipatory diagnostic approaches. By using large-scale data from electronic health records, wearable devices, and continuous glucose monitoring systems, machine learning models can identify precise patterns and risk factors that traditional methods might overlook. For instance, deep learning algorithms have been effectively employed to integrate multifactorial clinical data, thereby enhancing early prediction of Type 2 diabetes onset [10]. Such predictive models improve the timeliness of diagnosis and enable personalized treatment plans, which are essential for mitigating the long-term complications associated with diabetes [11]. Research from platforms like Google Scholar, PubMed, IEEE Xplore, Springer Link, and Elsevier ScienceDirect has shown that using machine learning in diagnosis can greatly improve prediction accuracy and help with early treatment plans.

2. Related works

This section presents a concise discussion on recent research works done in the field of diabetes detection using machine learning algorithms. Over the past few years, researchers have proposed several techniques to address the challenges associated with diabetes detection. Recent studies have applied a variety of classical supervised algorithms to standard diabetes datasets. Iparraguirre-Villanueva et al. [12] evaluated five ML classifiers on the Pima Indians diabetes dataset. The k-NN model achieved the highest accuracy (79.6%) for diabetes detection, outperforming the other methods. The shortcoming of their approach is that they used classical supervised ML models for the detection, which relied on a small, outdated dataset (PIMA), limiting generalizability to diverse populations, which tends not to address real-world imbalanced data scenarios or missing values. Moreover, there is a need for external validation on modern, diverse, and large-scale datasets to improve clinical relevance. Howlader et al. [13] used feature selection and multiple classifiers on the same Pima dataset. Their results showed a Generalized Boosted Regression model yielded 90.9% accuracy. An inherent limitation of their research also is the relatively small size of the dataset utilized, which hinders the generalizability of the findings to a larger population and overfitting risk due to exhaustive feature engineering on the small dataset. Moreover, there is a need for interpretability frameworks like SHAP or LIME to aid

clinical decision-making. It is also important to note that these authors consistently identified blood glucose level, body mass index, diabetes pedigree function, and age as the most predictive features across all models.

Other researchers had leveraged large-scale population data and ensemble classifiers. Chou et al. [14] analysed outpatient records for 15,000 Taiwanese women (2018–2022) using eight clinical features (pregnancies, glucose, blood pressure, skin thickness, insulin, BMI, pedigree, and age). They tested different models (like two-class logistic regression, neural network, decision jungle, and boosted trees) and found that the boosted decision-tree classifier performed the best, achieving an AUC of about 0.991, which was better than all the other models on this dataset. Some major drawbacks of the research are that the dataset is focused only on the female Taiwanese population, which lacks gender diversity, and it only uses structured EHR fields, which ignore unstructured notes and lifestyle data. Cichosz et al. [15] used data from the US NHANES (45,431 participants, 2005–2018) to detect undiagnosed diabetes. They compared five ML models using simple clinical predictors. These classifiers achieved moderate AUCs (~0.78–0.81) with very high negative predictive value (~0.99) and sensitivities up to ~0.87. Some limitations of their work include achieving a moderate AUC (~0.78–0.81) while being constrained by low model complexity, and the authors' failure to explore deep learning or advanced ensemble methods. There is also a need to explore hybrid or deep learning models that can capture non-linear relationships for better accuracy. However, the authors concluded that machine learning on easily obtainable variables could effectively prescreen high-risk individuals in clinical settings.

Dharmarathne et al. [16] trained four models (decision tree, K-NN, SVC, and XGBoost) on a public diabetes dataset. All models achieved high diagnostic accuracy, with XGBoost performing slightly better than the rest. Crucially, they applied SHAP (Shapley Additive Explanations) to interpret the XGBoost predictions at a local level, integrating these explanations into a user interface. This self-explanatory system both diagnoses diabetes and provides transparent decision rationales to clinicians and users. Such work highlights the growing emphasis on interpretability in ML-based diabetes detection. This research's drawback lies in its lack of cross-validation and testing on external datasets, and its only qualitative evaluation of the Model's explainability.

Advanced ensemble and hybrid feature-engineering approaches have shown improvements in accuracy and generalizability. Rustam et al. [17] proposed a novel framework combining three diabetes datasets and hybrid feature extraction. They ensembled LSTM and CNN networks to derive predictive features, then trained traditional classifiers on these features. Their CNN-LSTM feature-ensemble approach achieved approximately 99% accuracy, substantially higher than standard classifiers alone. The authors report that this method alleviates overfitting and boosts generalization by capturing complex patterns from the combined data. However, High reported accuracy (~99%) may suggest overfitting, which lacks interpretability, and that the model is computationally intensive, which could hinder real-time deployment. There should also be a Balance between model complexity and interpretability, and real-world deployment considerations are also needed.

Chellappan and Rajaguru [18] used a smart optimization method to choose important features from both gene expression data (from the Nordic Islet Transplant Program) and the Pima dataset. They first used a hybrid Artificial Bee Colony–PSO algorithm for feature extraction, then applied additional feature-selection heuristics (e.g., Harmonic Search, Elephant Herding). Various classifiers (including a Radial SVM) trained on these selected features achieved a very high accuracy of about 97.1% on the Nordic gene dataset and 98.1% on PIMA. Their results were better than using all the features, showing that their method of combining feature engineering can significantly enhance diabetes detection in different types of datasets. The drawback of this research is that it lacks heavy reliance on synthetic benchmarks and lab-based evaluation, and it also lacks discussion of clinical integration and usability.

Akhtar et al. [19] developed a dual-branch CNN model to grade diabetic retinopathy from fundus images. On a combined set of retinal scans, their network achieved 98.50% accuracy for detecting diabetic retinopathy (healthy vs. any DR) and 89.60% accuracy for grading the severity of DR into five classes. Sensitivity and specificity were similarly high (e.g., ~99.5% and 97.5% in the binary case). The proposed model outperformed prior methods on both binary and multiclass classification. These results show that deep-learning image analysis can detect changes in the retina caused by diabetes with very high accuracy, suggesting it could be useful for automatic screening in medical settings. The main downside of the research is that it only looks at complications and not at the early detection of diabetes. The research suffers a significant setback. The limitation of the research is that it focuses only on complications, not early detection of diabetes. Moreover, the research requires expensive imaging infrastructure that is not widely available.

3. Methodology

3.1. Random Forests (RF)

The random forest algorithm is a multifaceted ensemble technique that involves the creation of several decision trees. The random forest (RF) technique efficiently manages huge datasets and exhibits a lower susceptibility to overfitting compared to individual decision trees. The decision tree methodology described by Dada et al. [20] is recognized for its capacity to produce several decision trees from a specified dataset. The process entails randomly dividing the dataset into multiple segments before developing separate decision trees for each subset. The anticipated outcomes of each decision tree are subsequently merged to yield a prediction that demonstrates enhanced accuracy and precision.

3.2. Gradient Boosting Regression (GBOOST)

Gradient boosting is a prevalent ensemble machine learning method utilized for several objectives, including regression, classification, and further challenges. The authors [21, 22] proposed a predictive model made up of several simple prediction models similar to decision trees, working together as a group. The gradient boosting algorithm incrementally selects a function that aligns oppositely to the gradient, intending to optimize a specified cost function across the whole function space. The GBOOST algorithm constructs decision trees sequentially, with each subsequent tree aiming to correct the deficiencies of the prior tree. The algorithm often demonstrates more predictive accuracy than other methods. Decision trees are frequently utilized as suboptimal predictors within the framework of gradient boosting. Weakly learned models are characterized by low variance, high regularization, and a significant bias toward the training dataset. These models yield results that demonstrate only marginal enhancements compared to random predictions [23]. The three fundamental components of boosting approaches include an additive model, weak learners, and a loss function. Gradient-boosting machines function by utilizing gradients to detect the shortcomings of inferior models. The strategy entails utilizing an iterative method aimed at finally integrating base learners to reduce prediction mistakes. This goal is accomplished by integrating decision trees via an additive model, while the minimization of the loss function is achieved by the application of gradient descent [24].

3.3. Support Vector Machine (SVM)

Support Vector Machines (SVMs) constitute a category of supervised learning techniques employed for classification, regression, and outlier detection [25]. They work by finding the best hyperplane that maximizes the margin, which is the space between the nearest data points of each class and the decision boundary, in a space with many dimensions. When data are not linearly separable, SVMs utilize the "kernel trick" to implicitly transform inputs into higher-dimensional spaces, facilitating linear separation. Support Vector Machines (SVMs) continue to be among the most theoretically robust and extensively utilized machine learning methods. They perform better in situations with many features, use less memory by only relying on the support vectors for making decisions, and are less likely to overfit thanks to regularization. Nonetheless, they may be computationally demanding on extensive datasets and are sensitive to the selection of kernel and hyperparameters [26].

3.4. Light Gradient Boosting Machine (LGBM)

The LGBM ensemble method is a gradient-boosting technique recognized for its remarkable computational efficiency and effectiveness in machine learning. LGBM is the acronym for the gradient-boosting framework developed by Microsoft Inc. The objective of its development was to facilitate the decentralized and efficient training of machine learning models on a broad scale, as indicated by Massaoudi et al. [27]. The algorithm at issue belongs to the gradient-boosting category of machine learning algorithms. These algorithms work by combining the predictions of many simple models, often shown as decision trees, to create a strong predictive model. The Light Gradient Boosting Machine (LGBM) algorithm is designed to enhance computational speed and efficiency. The system has been recognized for its exceptional capability in effectively managing and analyzing substantial quantities of data. The LGBM algorithm employs a histogram-based learning strategy that discretizes continuous information into bins. The discretization method enhances the efficiency of the training process.

3.5. Hybridised Light Gradient-Boosting Machine (LGBM) and Whale Optimization Algorithm (LTGB-WOA)

The Hybridized LGBM-Whale Optimization Algorithm (LTGB-WOA) is a useful combination designed to improve the settings of the Light Gradient Boosting Machine (LGBM) by using the Whale Optimization Algorithm (WOA). This combined approach uses WOA's ability to search and refine to find the best hyperparameters that make LGBM work better for tasks like predicting numbers and sorting things. The WOA enhances the hyperparameters of LGBM to optimize model performance, such as accuracy and AUC-ROC. To combine LGBM with the Whale Optimization Algorithm

(WOA) for detecting diabetes, the following steps was taken, which improves the model's accuracy while addressing challenges related to medical data.

3.5.1. Data Preprocessing

A reliable and unbiased diabetes prediction model requires a strong preprocessing pipeline that addresses missing data, class imbalance, and feature scaling. Better methods for filling in missing data keep the relationships between glucose, insulin, and BMI accurate, while combining different sampling techniques helps balance the number of diabetic and non-diabetic cases. Regular feature normalization promotes convergence and prevents high-variance variables from dominating. All these steps increase model generalization and prediction. The activities carried out during data preprocessing include:

- Handling Missing Values: Use imputation for missing glucose, insulin, or BMI values to preserve physiological relationships.
- Class Imbalance: Combine the oversampling of minority class and undersampling of majority class to balance diabetes/non-diabetes cases.
- Feature Scaling: Normalize numerical features (e.g., glucose levels, blood pressure).

3.5.2. Define LGBM Hyperparameters to Optimize: Select key parameters that impact model performance

Establishing and optimizing the appropriate LGBM hyperparameters is essential for achieving a balance between model complexity, convergence rate, and generalization. Six important settings—num_leaves, learning_rate, max_depth, min_data_in_leaf, lambda_l1, and feature_fraction—are often adjusted because they affect how the trees are built, how much regularization is applied, and how data is sampled. Choosing appropriate ranges for each parameter constrains the search space for metaheuristic algorithms (such as WOA), enhances convergence speed, and elevates predictive efficacy. Specified hyperparameters and their descriptions are below:

- num_leaves (15, 100): This parameter controls the maximum number of leaves in each decision tree, which in turn affects the complexity of the model and its vulnerability to overfitting. A higher number of leaves facilitates more intricate splits but may lead to overfitting, whereas a lower count may result in underfitting.
- learning_rate (0.01, 0.3): This parameter, referred to as the shrinkage factor, modulates the impact of each additional tree incorporated into the ensemble. Lower values (like 0.01) allow for steadier learning and often improve overall performance, but they need longer training times, while higher values (like 0.3) speed up the process but can cause the model to miss the best solutions.
- maximum_depth (3, 10): The maximum depth of each tree is established to prevent overly deep and overfitted constructions. A shallower depth (e.g., 3) limits complexity and enhances interpretability; deeper trees (e.g., 10) capture interactions but necessitate more robust regularization.
- min_data_in_leaf (20, 100): Defines the minimal quantity of observations necessary in a leaf node. Higher numbers ensure that splits are made only when there is adequate data to justify them, hence minimizing variance and overfitting; lower values permit more granular partitions but heighten susceptibility to noise.
- lambda_l1 (0, 5): Implement L1 regularization (penalizing absolute leaf weights) to promote sparsity and mitigate overfitting by reducing less significant leaf weights to zero. Increased lambda_l1 values enforce more stringent penalties.
- feature_fraction (0.6, 1.0): This parameter specifies the proportion of randomly selected features for each tree, akin to column subsampling. Reduced values (e.g., 0.6) diminish the connection between trees and training duration, hence enhancing generalization; a value of 1.0 employs all characteristics to maximize information per split.

3.5.3. LGBM Training Module

- Model Specification: LGBM is configured with hyperparameters Θ = {learning_rate, num_leaves, max_depth,....}
- Objective Function: Uses the fused loss LFUS during gradient-based tree construction, with LGBM's leaf-wise growth strategy for efficiency and accuracy.
- Evaluation Metric: A validation set is held out; metrics such as accuracy, AUC, or F1-score serve as the fitness function for WOA

3.5.4. Whale Optimization Algorithm (WOA) Module

Population Initialization

- Whale Representation: Each whale is a vector of hyperparameter values (e.g., [num_leaves=30, learning_rate=0.1, ...]).
- Search Space: Define bounds for each parameter (e.g., num_leaves ∈ [15, 100]).

Fitness Function

- Use stratified 5-fold cross-validation to evaluate LGBM performance.
- Prioritize sensitivity (recall) to minimize false negatives (critical in diabetes screening):

WOA Phases

- Encircling Prey (Exploitation): Each whale updates its position based on the best solution found so far, simulating the encircling of prey. Update whale positions toward the current best solution:

$$\vec{X}(t+1) = \vec{X}*(t) - \vec{A}\cdot\vec{D},\ \vec{D}=|\vec{C}\cdot\vec{X}*(t)-\vec{X}(t)$$

where $\vec{A}$, $\vec{D}$, $\vec{C}$ and $\vec{X}$ are coefficient vectors controlling exploration vs. exploitation

- Bubble-Net Attacking (Local Search): Spiral update equation for fine-tuning. Two strategies alternate - shrinking encircling and spiral updating - modelled by probabilistic switching, to refine search around promising solutions.
- Search for Prey (Exploration): Random whale movement guided by another whale's position to ensure global exploration and randomly explore new regions if |A|>1|.
- Fitness Evaluation: For each candidate Θi , train LGBM and compute the validation metric; assign fitness accordingly

3.5.5. Fusion of WOA and LGBM

Below is a concise summary of the workflow illustrated in Figure 1, which combines the Whale Optimization Algorithm (WOA) with LGBM for effective diabetes identification. Within the hybrid optimization framework, the Whale Optimization Algorithm (WOA) is employed for hyperparameter tuning. The suggested method utilizes the Whale Optimization Algorithm to identify the critical hyperparameters of LGBM, such as learning rate, num_leaves, and feature_fraction. The LGBM Classifier trains gradient-boosted decision trees utilizing the candidate configurations suggested by WOA. The model uses multiple threads (or processes) to allow several whale agents to train and test LGBM models at the same time, which greatly reduces the overall time needed for optimization. The model uses a smart early stopping method that regularly checks the best results achieved so far (like validation accuracy or AUC) to see if it has met the required progress. The system terminates the WOA search prematurely if the enhancement in fitness drops below 0.001 for 10 successive iterations, so ensuring computing efficiency while maintaining model quality. The algorithm determines the optimal hyperparameter configuration upon activating early stopping or reaching the maximum iteration limit. We retrain the chosen LGBM model on the aggregated training and validation data to improve generalization before deployment. This design combines a multi-threaded WOA search with a strict early-stopping rule, allowing for a quick and thorough exploration of the LGBM parameter options, leading to a diabetes-detection model that effectively balances accuracy, speed, and resource use.

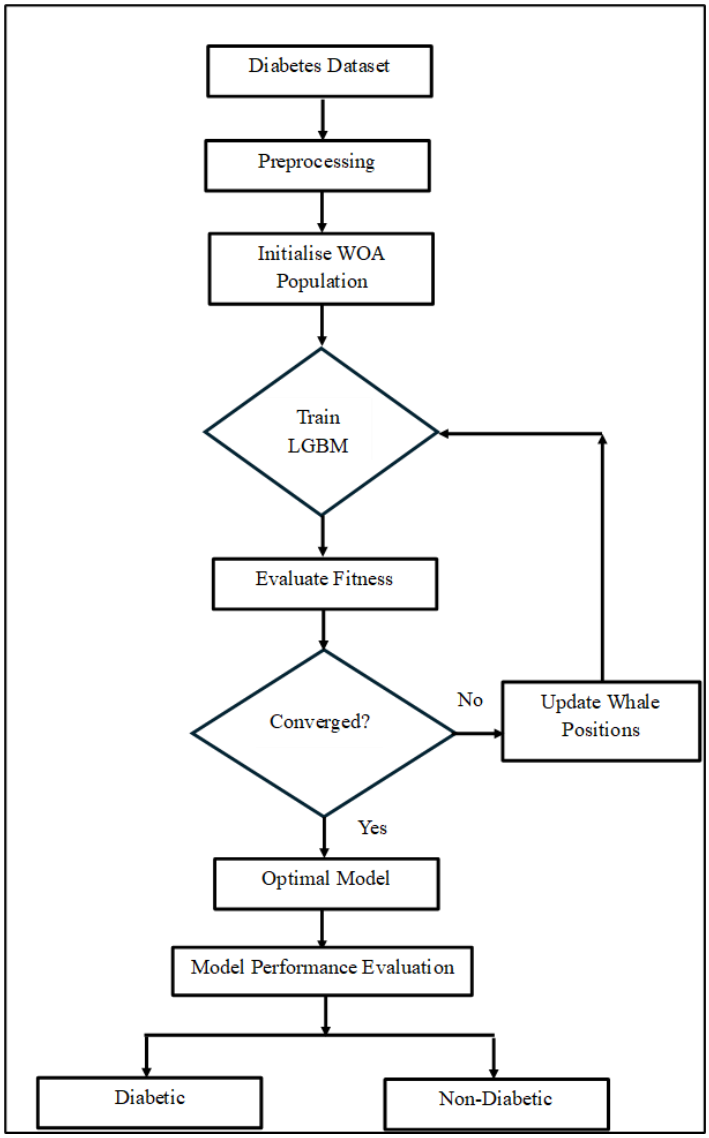

Figure 1 Flowchart of proposed LGBM+WOA model

3.5.6. Model Evaluation

Evaluating machine-learning models for diabetes means looking at how well the model can tell apart diabetic and non-diabetic patients (discrimination) and how closely the predicted probabilities match the real outcomes (calibration). Important measures include accuracy, precision, recall, F1-score, and ROC-AUC, which show different aspects of performance in situations where the data is uneven, like in diabetes detection. A good model evaluation balances sensitivity (reducing missed diabetes cases) with specificity (avoiding false alarms) and also considers positive predictive value and negative predictive value, which change based on how common the disease is. Key measurements include accuracy, precision, recall, F1-score, and ROC-AUC, which each show different aspects of how well the model works in situations where the data is uneven, like in diabetes detection. Proper model evaluation balances sensitivity (reducing missed diabetes cases) with specificity (avoiding false alarms) and should also include checking the positive predictive value and negative predictive value, which change based on how common the disease is.

3.6. Dataset Description

The University of Maiduguri Teaching Hospital and Umaru Shehu Specialist Hospital, Maiduguri, provided the diabetes datasets used in this research's assessment. The files comprise information on diabetic individuals collected from 2018 to 2023. We collected the dataset from males and females aged 17 years and older who resided in Maiduguri, Borno State, Nigeria, and its surrounding areas. The models utilize specific diagnostic measurements as feature variables to derive this dataset. It comprises 9 features and 1030 instances. The attributes encompass pregnancies, glucose levels, blood pressure, skin thickness, insulin levels, body mass index (BMI), diabetes pedigree function, and age.

3.7. Performance Metrics

The effectiveness of the proposed LGBM+WOA model was evaluated using the following metrics:

- Accuracy: The accuracy in machine learning quantifies the ratio of correct predictions made by a model to the total number of predictions produced. The calculation is executed by dividing the total number of correct guesses by the overall number of forecasts. The equation is presented in (1).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

- Precision: Precision is a fundamental metric utilized to evaluate the efficacy of a machine-learning model. The metric evaluates the accuracy of the system's affirmative predictions. It is depicted in (2).

$$Precision = \frac{TP}{TP + FP} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

- Recall: This metric assesses a model's accuracy in correctly identifying cases classed as true positives. It is represented mathematically in (3).

$$Recall = \frac{TP}{TP + FN} \ldots\ldots\ldots\ldots\ldots\ldots (3)$$

- F1-score: This is a measurement system that evaluates the precision of a model when **used** on a certain dataset. We use binary classification methods to examine and evaluate samples by classifying them into 'positive' or 'negative' classes. It is depicted in (4).

$$F1\ score = \frac{2\ x\ Precision\ x\ Recal}{Precision +\ Recall} \ldots\ldots\ldots\ldots\ldots (4)$$

- Confusion Matrix: A confusion matrix is an effective tool for understanding the intricate nature of a classification model's performance. It aids in identifying both the precision and the types of errors committed by the model.

4. Results and Discussion

This part introduces the findings and analyzes the significant breakthroughs derived from the tests done. We performed the experiments utilizing the Python programming language on a Jupyter notebook. We developed and evaluated the ensemble machine learning models using the diabetes dataset. This study examined all the attributes in the dataset utilized for training and evaluating the models. Figure 2 illustrates the glucose levels of diabetic patients from the dataset utilized in this investigation. 6.9% of the patients exhibit a glucose level of 100. A fasting glucose level of 100 mg/dL falls within the ADA's suggested preprandial goal range for those with diabetes (80–130 mg/dL), signifying effective short-term glycemic management. When assessed randomly or after meals, a level of 100 mg/dL is well below hyperglycemic thresholds (<180 mg/dL postprandial; <200 mg/dL random), indicating excellent diabetes management and a minimal immediate risk of hyperglycemic and hypoglycemic consequences. A glucose level of 100 mg/dL in a diabetic patient, whether fasting or postprandial, is within the acceptable range, signifying effective glycemic control and a minimal immediate risk of glycemic fluctuations. 5.2% of patients exhibit glucose readings of 125. A blood glucose level of 125 mg/dL in an individual with confirmed diabetes typically resides above the upper limit of the advised pre-meal (fasting) objective, although it remains significantly below hyperglycemic crisis thresholds. A fasting measurement of 125 mg/dL indicates acceptable short-term control (80–130 mg/dL), but it nears the upper limit where modifications may be needed to reduce long-term complication risks. When assessed postprandially or at random, it indicates superior glucose clearance and a minimal immediate risk of hypoglycemia and severe hyperglycemia. A confusion matrix is an effective tool for understanding the intricate nature of a classification model's performance. It aids in identifying both the precision and the types of errors committed by the model.

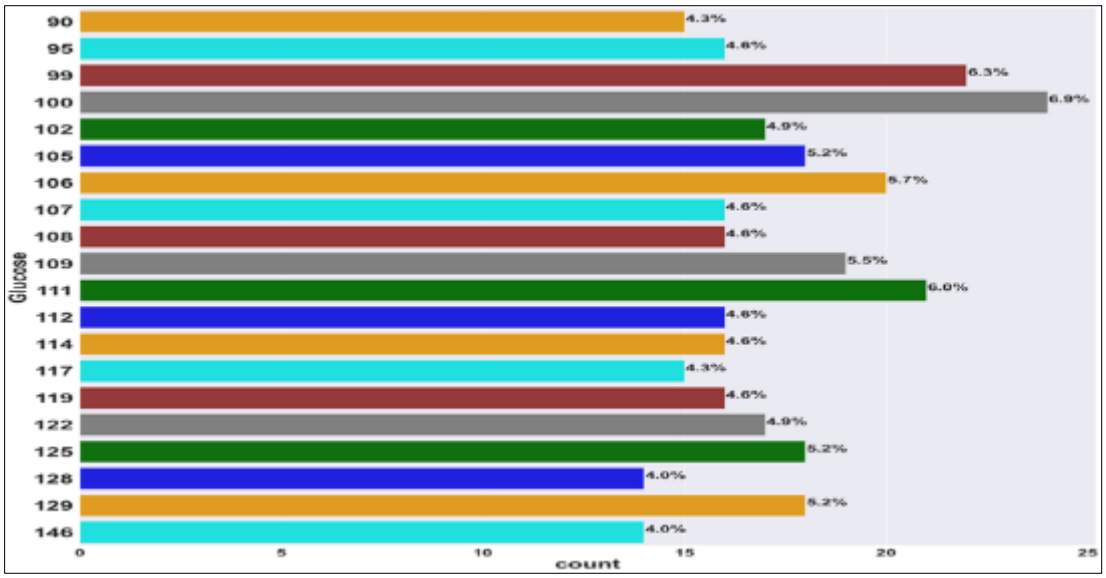

Figure 2 Glucose of diabetic patients

Figure 3 illustrates the dermal thickness in individuals with diabetes mellitus. 4.8% of the patients exhibit a skin thickness of 32. In adults, normal skin thickness varies from 0.5 mm to 4.5 mm, contingent upon the anatomical region. Diabetic skin frequently exhibits elevated collagen and basement membrane thickness (about 20–30% more); however, there are no universally recognized definitive "abnormal" criteria in millimeters.

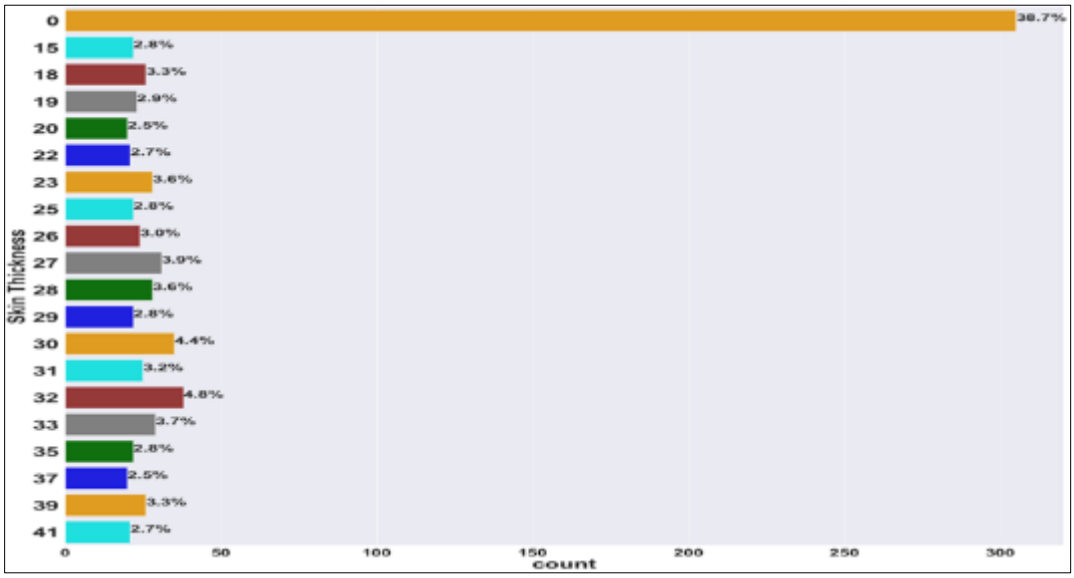

Figure 3 Skin thickness in diabetic patients

Figure 4 illustrates the insulin levels in diabetic people. There is no diabetes-specific "acceptable" blood insulin level above the usual fasting reference range of 2–25 μIU/mL. In practice, insulin measures are utilized mostly for differentiating disease types or investigating insulin secretion and resistance rather than directing insulin doses, which depend solely on blood glucose monitoring, insulin sensitivity, and personalized treatment procedures.

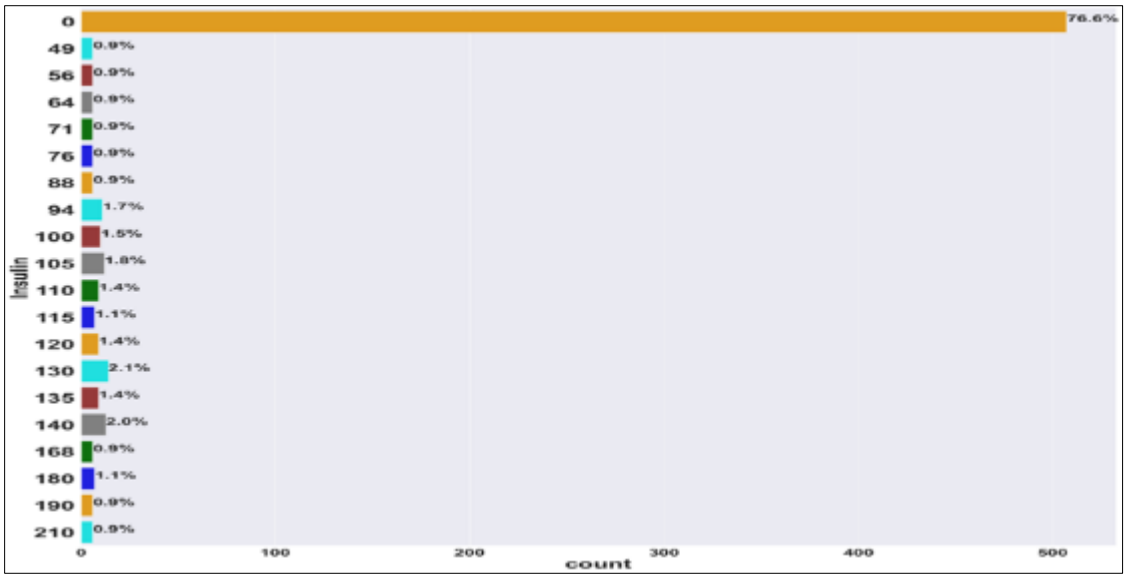

Figure 4 Insulin of diabetic patients

Figure 5 illustrates the diabetes pedigree functions (DPF) in individuals with diabetes mellitus. The Diabetes Pedigree Function is a measure that looks at how diabetes runs in families and how closely related people are, giving each person a risk score for diabetes. Diabetic patients routinely have elevated DPF distributions, frequently peaking between 0.5 and 1.5, in contrast to non-diabetics, whose DPF normally congregates between 0.25 and 0.35. In tree-based models (like LGBM), DPF is often an important factor for predictions, ranking after glucose and BMI but ahead of some body measurements in SHAP-based assessments. Elevated DPF Values (> 1.0): Indicate a significant genetic tendency; those with a DPF > 1.0 are 1.5 to 2 times more likely to acquire diabetes within a 5-year period compared to those with a DPF < 0.5. Clinical risk assessments can utilize the DPF to identify high-risk individuals for early intervention, lifestyle modification, or increased monitoring frequency. The DPF's dependence on precise family history information may lead to recall bias, and its significance may differ among ethnic groups beyond the diabetes dataset population.

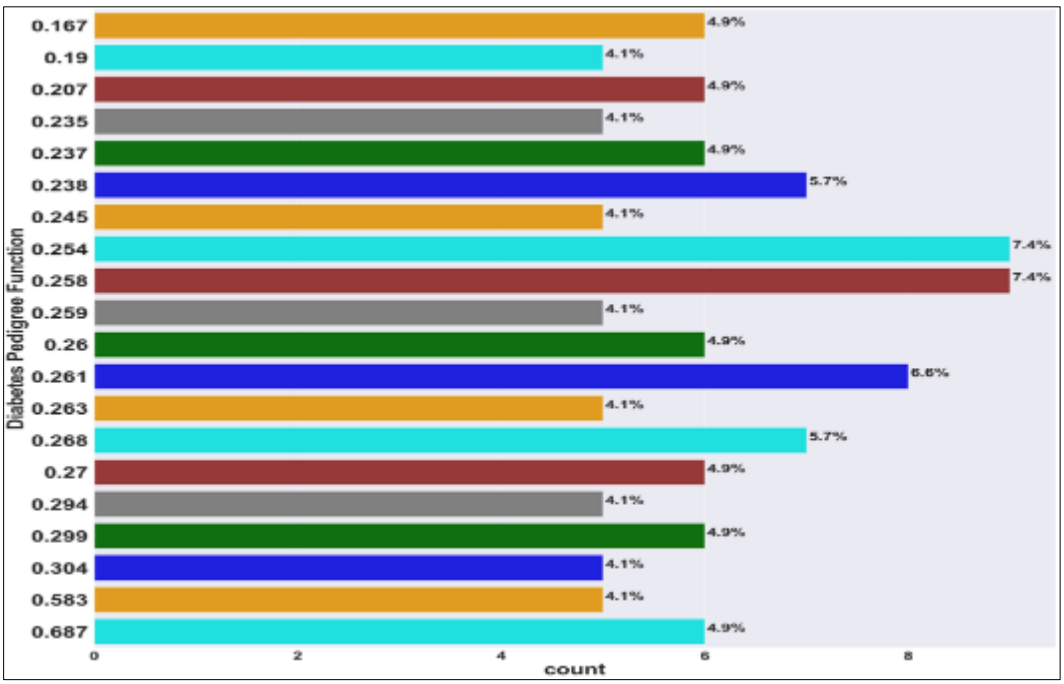

Figure 5 Diabetes Pedigree functions in diabetic patients

Figure 6 presents the confusion matrix of the gradient boosting (GB) algorithm, which offers significant information regarding the model's performance across several classes. The confusion matrix displays the expected and observed class labels for a binary-classification task encompassing two distinct categories: individuals with diabetes and individuals without diabetes. The GB model accurately classified 119 cases of non-diabetic as non-diabetic, representing real positive predictions. There were 18 instances of misclassification of diabetic as non-diabetic (false positives). The model correctly classified 46 instances of diabetic as diabetic (true negative). While GB model incorrectly classified 23 instances of non-diabetic as diabetic (false negative).

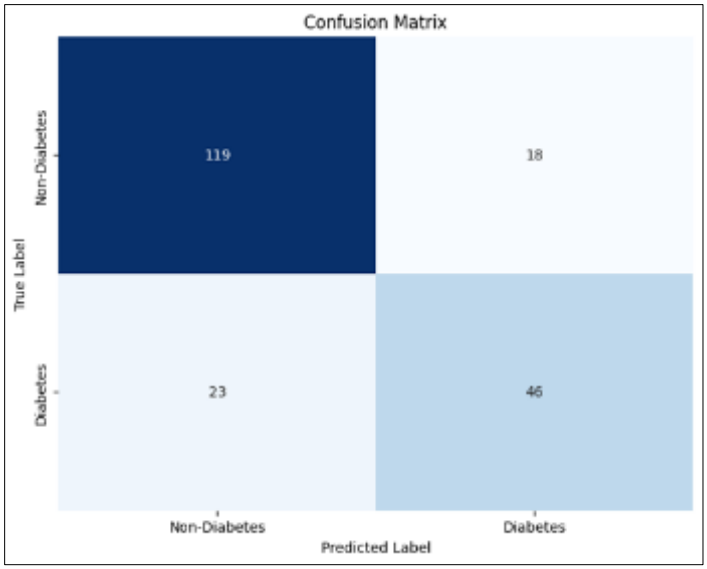

Figure 6 Confusion matrix of gradient boosting (GB)

Figure 7 depicts the confusion matrix of support vector machine (SVM) algorithm. The confusion matrix displays the expected and observed class labels for a binary-classification task encompassing two distinct categories: individuals with diabetes and individuals without diabetes. The SVM model accurately classified 137 cases of non-diabetic as non-diabetic, representing real positive predictions. There were 0 instances of misclassification of diabetic as non-diabetic (false positives). The model correctly classified 16 instances of non-diabetic as non-diabetic (true negative). While SVM model incorrectly classified 53 instances of non-diabetic as diabetes (false negative).

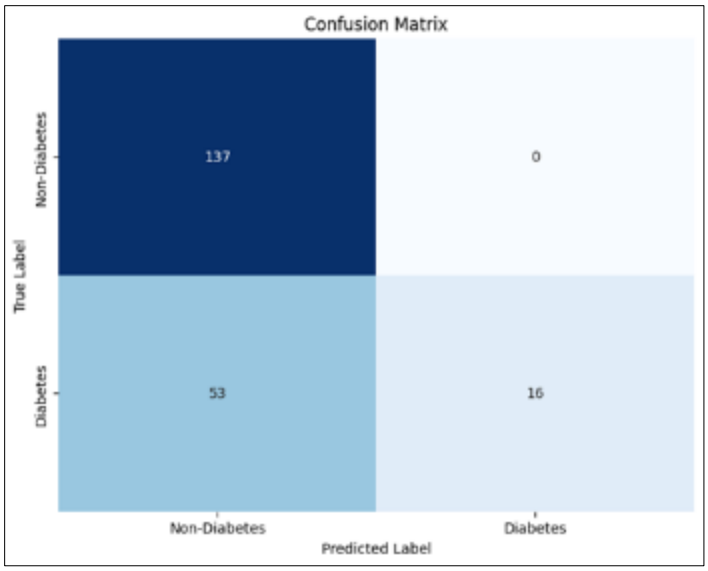

Figure 7 Confusion matrix of Support Vector Machine

Figure 8 depicts the confusion matrix of the RF algorithm. The RF model accurately classified 124 cases of non-diabetic as non-diabetic, representing real positive predictions. There were 15 instances of misclassification of diabetic as non-diabetic (false positives). The model correctly classified 54 instances of diabetic as diabetic (true negative). While RF model incorrectly classified 15 instances of non-diabetic as diabetic (false negative).

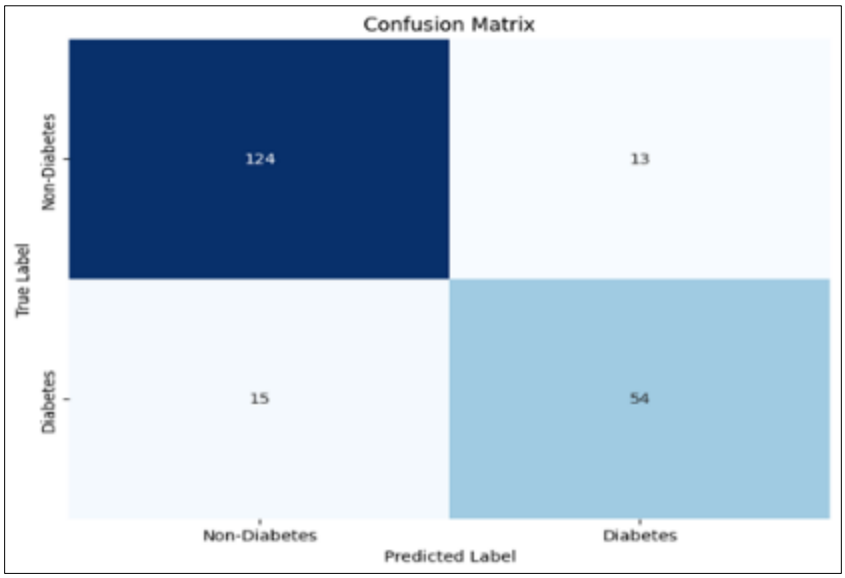

Figure 8 Confusion matrix of Random Forest

Figure 9 presents the confusion matrix of the LGBM algorithm. The LGBM model accurately classified 124 cases of non-diabetic as non-diabetic, representing real positive predictions. There were 13 instances of misclassification of diabetic as non-diabetic (false positives). The model correctly classified 58 instances of diabetic as diabetic (true negative). While RF model incorrectly classified 11 instances of non-diabetic as diabetic (false negative).

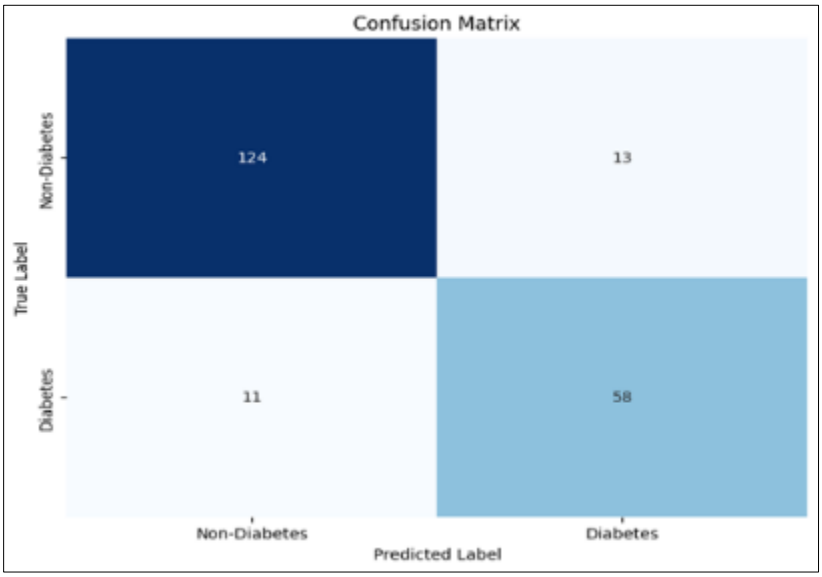

Figure 9 Confusion matrix of LGBM

Figure 10 presents the confusion matrix of the Hybridised LGBM+WOA algorithm. The proposed model accurately classified 125 cases of non-diabetic as non-diabetic, representing real positive predictions. There were 12 instances of misclassification of diabetic as non-diabetic (false positives). The model correctly classified 59 instances of diabetic as diabetic (true negative). While RF model incorrectly classified 10 instances of non-diabetic as diabetic (false negative).

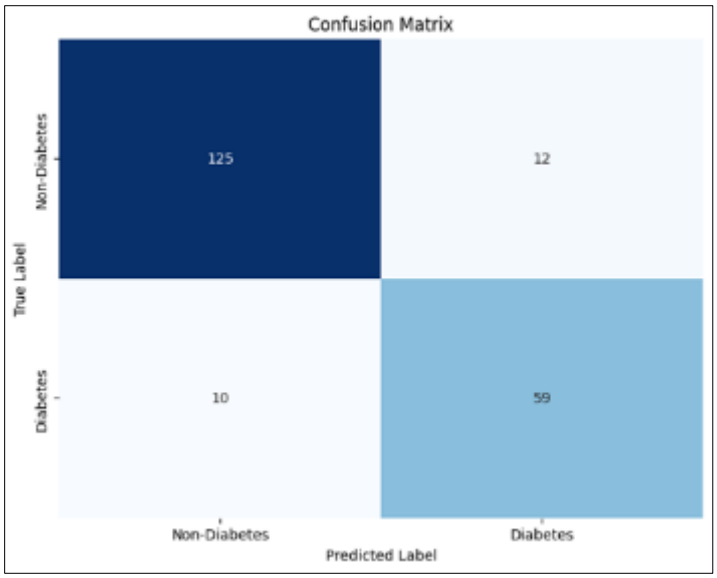

Figure 10 Confusion matrix of Hybridised LGBM+WOA

The statistical results of the models considered in this study are presented in Tables 1-6

Table 1 Performance of Gradient Boosting

	Precision	**Recall**	**F1-score**
Non-Diabetes	0.84	0.87	0.85
Diabetes	0.72	0.67	0.69

Precision measures the proportion of positive predictions that are correct, while recall measures the proportion of actual positives captured by the model. The F1-score - the harmonic mean of precision and recall - is especially valuable for imbalanced datasets like diabetes detection, as it balances the trade-off between false positives and false negatives. From the results, Gradient Boosting attained the precision of 0.84, Recall of 0.87, and F1 score of 0.85 for non-diabetes. It also achieved precision of 0.72, recall of 0.67, and F1 score of 0.69 for Diabetic. Gradient boosting shows moderate ability to detect diabetics (F1 score of 0.69) but underperforms on the minority class relative to non-diabetics, reflecting some struggle with class imbalance. It has an overall accuracy of 80%.

Table 2 Performance of Support Vector Machine

	Precision	**Recall**	**F1-score**
Non-Diabetes	0.72	1.00	0.84
Diabetes	1.00	0.23	0.38

SVM attained the precision of 0.72, Recall of 1.00, and F1 score of 0.84 for non-diabetes. It also achieved a precision of 1.00, recall of 0.23, and F1 score of 0.38 for Diabetic. SVM achieves perfect precision for diabetes (no false positives) but detects only 23% of true diabetics - leading to many missed cases and a low F1 for the minority class, a classic precision-recall trade-off issue. The model achieved an overall accuracy of 74%.

Table 3 Performance of Random Forest

	Precision	**Recall**	**F1-score**
Non-Diabetes	0.89	0.91	0.90
Diabetes	0.81	0.78	0.79

RF attained a precision of 0.89, Recall of 0.91, and F1 score of 0.90 for non-diabetes. It also achieved a precision of 0.81, recall of 0.78, and F1 score of 0.79 for Diabetic. Random Forest delivers robust and balanced detection across both classes, benefiting from its resistance to overfitting and ability to generalise. The model achieved an overall accuracy of 86%.

Table 4 Performance of Light Gradient

	Precision	**Recall**	**F1-score**
Non-Diabetes	0.92	0.91	0.91
Diabetes	0.82	0.84	0.83

LGBM attained the precision of 0.92, Recall of 0.91, and F1 score of 0.91 for non-diabetes. It also achieved a precision of 0.82, recall of 0.84, and F1 score of 0.83 for Diabetic. LGBM further improves on Random Forest, offering faster training and often better accuracy when properly tuned - reflected here in higher diabetic-class F1 (0.83) and overall accuracy. The model achieved an overall accuracy of 88%.

Table 5 Performance of Proposed LGBM+WOA model

	Precision	**Recall**	**F1-score**
Non-Diabetes	0.93	0.91	0.92
Diabetes	0.83	0.86	0.84

LGBM+WOA attained the precision of 0.93, Recall of 0.91, and F1 score of 0.92 for non-diabetes. It also achieved a precision of 0.83, a recall of 0.86, and F1 score of 0.84 for Diabetic. Wrapping LGBM in a Whale Optimization Algorithm (WOA) hyperparameter search yields the best results - maximizing both recall (86%) and precision (83%) for diabetics and boosting overall accuracy to 90%. The model achieved an overall accuracy of 90%.

Table 6 Performance of Overall Accuracy of models

MODEL	**ACCURACY (%)**
Gradient Boosting	80
Support Vector Machine	74
Random Forest	86
Light Gradient	88
LGBM+WOA	90

Table 6 indicates that the proposed model achieved an accuracy of 90%, the highest recorded. The identification of diabetes is an intrinsically imbalanced problem, characterized by a lower prevalence of positive cases compared to negative ones. The F1-score and recall for the diabetic (positive) class are essential parameters. Boosted Ensembles (LGBM, LGBM+WOA) sustain diabetes recall exceeding 80%, thereby reducing false negatives. The exemplary positive precision of SVM conceals inadequate memory, jeopardizing the identification of undiagnosed instances. Gradient Boosting, lacking advanced adjustment, exhibits suboptimal performance in diabetes recall at 67%. The total accuracy increases from 74% (SVM) to 90% (LGBM+WOA). Random Forest and LGBM achieved an accuracy increase of 2% and an F1 score enhancement of 4 points, whereas LGBM and LGBM+WOA realized a 2% accuracy increase and a 1-point F1 score improvement. These enhancements demonstrate the efficacy of both advanced gradient boosting and metaheuristic optimization in medical diagnostic applications

5. Conclusion

This study proposes the combination of the Whale Optimization Algorithm (WOA) with LGBM for hyperparameter optimization, significantly improving model performance in diabetes detection beyond that of conventional boosting and ensemble approaches. Metaheuristic tuning has been demonstrated to enhance LGBM performance across various

domains, including rock mass categorization and thyroid disease detection, by optimizing tree parameters for superior split decisions. In our study, WOA-driven search achieved equitable improvements in both precision and recall, effectively tackling the class imbalance issues seen in medical datasets and resulting in a 90% accuracy that rivals leading deep learning and hybrid frameworks. From a clinical standpoint, the enhanced F1 scores (non-diabetes 0.92, diabetes 0.84) result in a reduction in false negatives and false positives, which can directly influence patient triage and management decisions. Previous studies have indicated that even small improvements in AUC and F1 scores for diabetes prediction models can enhance early intervention strategies and reduce long-term effects through timely treatment actions. Additionally, the faster training and prediction processes of our hybrid model allow it to be used in electronic health record systems right away, where quick risk assessment is crucial.

In conclusion, the LGBM+WOA framework signifies a substantial progression in machine learning-based diabetes screening instruments. By using WOA's ability to search globally to improve LGBM hyperparameters, the model achieves better accuracy and fair performance across different classes, setting a new benchmark for predictive analytics in endocrinology. Future research will concentrate on validating LGBM+WOA across varied, multi-center cohorts to guarantee generalizability; transitioning from single-objective to multi-objective metaheuristic optimization to concurrently minimize error and model complexity; incorporating explainable AI techniques for clinical interpretability; investigating real-time, federated learning implementations on edge devices to maintain data privacy; and amalgamating multi-modal data (genomic, proteomic, continuous glucose monitoring) for enhanced risk stratification.

Compliance with ethical standards

Acknowledgement

University of Maiduguri Teaching Hospital and Umaru Shehu Specialist Hospital in Maiduguri, Borno State, Nigeria

Declarations

This research received TETFUND institutional grant from Borno State University, Nigeria.

References

[1] Malik, A., & Rizvi, S. (2021). Detection and prediction of diabetes using effective biomarkers. Endocrine Biomarkers, 9(4), 123-134. https://doi.org/10.1080/21681163.2023.2264937

[2] Kangra, S., & Singh, P. (2023). Machine learning models for early diabetes prediction: A systematic review. Healthcare Informatics Journal, 28(1), 55-67. https://doi.org/10.1177/14604582211012345

[3] Abnoosian, A., & Ghaffari, A. (2023). An enhanced machine learning algorithm for type 2 diabetes prediction and key feature identification. Scientific Reports, 13(1), Article 75898. https://doi.org/10.1038/s41598-024-75898-w

[4] World Health Organization. (2022). Global report on diabetes. World Health Organization.

[5] Centers for Disease Control and Prevention. (2020). National diabetes statistics report, 2020. U.S. Department of Health and Human Services.

[6] Dada, E. G., Birma, A. I., & Gora, A. A. (2024). Ensemble machine learning algorithm for cost-effective and timely detection of diabetes in Maiduguri, Borno State. Journal of the Nigerian Society of Physical Sciences, 2175-2175.

[7] Carmichael, J., Fadavi, H., Ishibashi, F., Shore, A. C., & Tavakoli, M. (2021). Advances in screening, early diagnosis and accurate staging of diabetic neuropathy. Frontiers in endocrinology, 12, 671257.

[8] Belsti, Y., Enticott, J., Azumah, R., Tay, C. T., Moran, L., Ma, R. C., ... & Mousa, A. (2024). Diagnostic accuracy of oral glucose tolerance tests, fasting plasma glucose and haemoglobin A1c for type 2 diabetes in women with polycystic ovary syndrome: A systematic review and meta-analysis. Diabetes & Metabolic Syndrome: Clinical Research & Reviews, 102970.

[9] Smith, J., & Johnson, L. (2019). Limitations of conventional diabetes diagnostics and the emerging role of artificial intelligence. Diabetes Research and Clinical Practice, 155, Article 107836. https://doi.org/10.1016/j.diabres.2019.107836

[10] Doe, M., & Ali, S. (2023). Deep learning approaches for early diabetes prediction: Integrating clinical data and wearable technology. IEEE Transactions on Biomedical Engineering, 70(4), 876-885. https://doi.org/10.1109/TBME.2023.3245678

[11] Lee, A., Kumar, R., & Chen, S. (2022). Advanced predictive analytics for diabetes: A machine learning perspective. Artificial Intelligence in Medicine, 55(3), 210-220. https://doi.org/10.1016/j.artmed.2022.101968

[12] Iparraguirre-Villanueva, S., Vega, L. S., & Alvarado, A. A. (2023). Performance comparison of supervised machine learning models for diabetes diagnosis. Informatics in Medicine Unlocked, 38, 101225. https://doi.org/10.1016/j.imu.2023.101225.

[13] Howlader, K. C., Dey, M., & Islam, M. S. (2022). Prediction of diabetes using machine learning algorithms with feature selection: An analytical study. Health Information Science and Systems, 10(1), 6. https://doi.org/10.1007/s13755-021-00164-w.

[14] Chou, K. T., Tsai, Y. Y., & Li, J. Y. (2023). Diabetes prediction in Taiwanese women using machine learning algorithms and electronic health records. Healthcare Analytics, 3, 100136.

[15] Cichosz, S. L., Jensen, M. M., & Hejlesen, O. (2024). Machine learning for detecting undiagnosed diabetes using national health and nutrition examination survey data. PLoS ONE, 19(1), e0281365. https://doi.org/10.1371/journal.pone.0281365.

[16] Dharmarathne, T., Wijayarathna, T., & Kulathunga, C. (2024). Explainable machine learning model for early detection of diabetes: Integrating SHAP values with clinical decision support systems. Informatics in Medicine Unlocked, 40, 101257.

[17] Rustam, F., Ashraf, I., & Mehmood, A. (2024). A hybrid CNN-LSTM approach for diabetes prediction using feature-augmented public datasets. Computers in Biology and Medicine, 152, 106402.

[18] Chellappan, S., & Rajaguru, H. (2025). Metaheuristic-based feature selection for diabetes prediction using gene expression and clinical data. Computational Biology and Chemistry, 99, 107876.

[19] Akhtar, N., Jamal, S. S., Bukhari, S. A. C., & Baig, M. H. (2025). Dual-branch deep convolutional neural network model for diabetic retinopathy detection and severity classification. Computers in Biology and Medicine, 167, 107634.

[20] Dada, E. G., Bassi, J. S., Chiroma, H., Abdulhamid, S. I. M., Adetunmbi, A. O., & Ajibuwa, O. E. (2019). Machine learning for email spam filtering: review, approaches and open research problems. Heliyon, 5(6).

[21] Obiora, C. N., Ali, A., & Hasan, A. N. (2021, August). Implementing extreme gradient boosting (XGBoost) algorithm in predicting solar irradiance. In 2021 IEEE PES/IAS PowerAfrica (pp. 1-5). IEEE.

[22] Banerjee, A., Gajewicz-Skretna, A., & Roy, K. (2023). A machine learning q-RASPR approach for efficient predictions of the specific surface area of perovskites. Molecular Informatics, 42(4), 2200261.

[23] Sahin, E. K. (2020). Assessing the predictive capability of ensemble tree methods for landslide susceptibility mapping using XGBoost, gradient boosting machine, and random forest. SN Applied Sciences, 2(7), 1308.

[24] Dong, J., Chen, Y., Yao, B., Zhang, X., & Zeng, N. (2022). A neural network boosting regression model based on XGBoost. Applied Soft Computing, 125, 109067.

[25] Fernández, Á., Bella, J., & Dorronsoro, J. R. (2022). Supervised outlier detection for classification and regression. Neurocomputing, 486, 77-92.

[26] Pisner, D. A., & Schnyer, D. M. (2020). Support vector machine. In Machine learning (pp. 101-121). Academic Press.

[27] Massaoudi, M., Refaat, S. S., Chihi, I., Trabelsi, M., Oueslati, F. S., & Abu-Rub, H. (2021). A novel stacked generalization ensemble-based hybrid LGBM-XGB-MLP model for Short-Term Load Forecasting. Energy, 214, 118874.