

OPTIMIZATION OF DEEP NEURAL NETWORKS FOR HEART DISEASE DIAGNOSIS USING THE AQUILA OPTIMIZER: BRIDGING AI AND BIO-INSPIRED COMPUTATION

Agu Edward Onyebueke (PhD)^{1*}, Oladunjoye John Abiodun (PhD)², Andrew Ishaku Wreford³

^{1,2,3} Computer Science Dept Federal University Wukari, Taraba State, Nigeria.

*Corresponding Author

Agu Edward Onyebueke
(PhD)

Computer Science Dept
Federal University Wukari,
Taraba State, Nigeria.

Article History

Received: 19.10.2025
Accepted: 29.12.2025
Published: 23.01.2026

Abstract: Cardiovascular diseases (CVDs) are still the primary cause of death globally, and heart disease is the most critical contributor. Precise diagnosis early on prevents poor outcomes, but available diagnosis techniques are typically costly, invasive, and dependent on expert interpretation. This study explores whether artificial intelligence (AI), in the guise of deep learning and bio-inspired optimization, can provide a robust, non-invasive predictive diagnosis of heart disease. Two baseline algorithms, Multi-Layer Perceptron (MLP) and Artificial Neural Network (ANN), were trained and tested on the Cleveland heart disease database. While baseline performance was found to have outstanding predictive capability, this was constrained by sensitivity to hyperparameters. To alleviate this, tuning was performed using the new bio-inspired metaheuristic, the Aquila Optimizer (AO), which draws inspiration from the hunting strategy of Aquila eagles. Experimental results show that AO significantly enhances model performance: the AO-tuned MLP achieved accuracy of 98.05% and ideal precision of 100%, while the AO-tuned ANN produced optimal performance with 98.54% accuracy, 100% precision, 97.09% recall, and a Matthews Correlation Coefficient (MCC) of 0.9712. Convergence of the fitness function confirmed AO's capability to achieve stable and near-optimum solutions in few iterations. Results emphasize the benefit of ANNs for structured clinical data and validate that AO is a phenomenal tool for optimizing neural networks in medicine.

Keywords: Heart Disease Prediction, Artificial Neural Network (ANN), Multi-Layer Perceptron (MLP), Aquila Optimizer (AO), Hyperparameter Tuning.

Cite this article:

Onyebueke, A. E., Abiodun, O. J., Wreford, A. I., (2026). OPTIMIZATION OF DEEP NEURAL NETWORKS FOR HEART DISEASE DIAGNOSIS USING THE AQUILA OPTIMIZER: BRIDGING AI AND BIO-INSPIRED COMPUTATION. *ISAR Journal of Science and Technology*, 4(1), 62-70.

1.0 Introduction

Cardiovascular diseases (CVDs) remain the globe's leading cause of mortality, taking around 17.9 million lives annually (WHO, 2023) [1]. Heart disease is the most prominent burden among them, found in developed as well as developing nations, with increasing prevalence due to sedentary lifestyles, poor diet, and a growing geriatric population. Prompt diagnosis of heart disease is vital since early treatment can significantly reduce the risk of myocardial infarction, stroke, and cardiac arrest. Invasive conventional diagnostic procedures such as electrocardiography (ECG), angiography, echocardiography, and treadmill stress tests although accurate are typically costly, invasive, time consuming, and prone to specialist interpretation [2]. Answering these challenges,

artificial intelligence (AI) has become a more sophisticated vehicle of surmounting them through the provision of automated, data-derived prediction models. AI is able to read patient histories, laboratory tests, and imaging studies in order to detect concealed patterns that are hard or impossible for specialists to detect individually, thus guiding clinical decision-making and enabling preventive interventions [3].

In the past two decades, various paradigms of AI have been used in heart disease databases. Machine Learning (ML) algorithms like Decision Trees, Random Forests, Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Logistic Regression have shown robust baseline performance in heart disease prediction [4, 5]. These approaches work well in datasets with structured, moderate-



dimension data but begin to struggle when confronted with highly non-linear feature interactions. Ensemble Learning approaches such as Gradient Boosting, AdaBoost, and XGBoost have harnessed the strength of multiple weak classifiers to improve predictive accuracy [6]. These approaches have been able to deliver good accuracy but often at the cost of interpretability, which is one of the main requirements in clinics. Deep Learning (DL) networks, specifically Artificial Neural Networks (ANNs), Multi-Layer Perceptrons (MLPs), Convolutional Neural Networks (CNNs), and Recurrent Neural Networks (RNNs), have shown exceptional ability to extract high-level feature hierarchies [7]. For structured clinical data, MLPs and ANNs are especially relevant because they can best represent interactions between patient risk factors such as cholesterol, blood pressure, age, and chest pain type. The RNNs and CNNs, although widely employed in time-series processing and healthcare imaging, prove less effective in small table-based datasets, but their success represents the broader potential of AI in medicine. The different approaches highlight the potential of AI to predict heart disease but also highlight issues such as overfitting, computational cost, hyperparameter sensitivity, and class imbalance. ANNs remain at the center of AI-driven heart disease diagnosis due to their ability to estimate non-linear functions and find hidden relationships among patient attributes. The standard form of ANN is MLP, which consists of an input layer, multiple hidden layers, and an output layer. It is the best-established architecture for formal medical data with a good balance between complexity and interpretability.

However, the prediction performance of ANNs and MLPs is highly dependent on hyperparameters such as the number of neurons in a hidden layer, learning rate, batch size, and dropout rate [8]. Hence, poor selection of these parameters leads to sub-optimal convergence, reduced accuracy, or poor generalization on unknown patient records. Such sensitivity makes it a priority that effective hyperparameter tuning techniques be employed to unlock the maximum potential of ANNs and MLPs for clinical prediction tasks.

Classical hyperparameter tuning methods such as grid search and random search are computationally costly and susceptible to getting stuck at a local optimum [9]. In opposition to these, bio-inspired optimization algorithms based on biological functions and processes have become alternative solutions of enormous potential.

Algorithms such as Genetic Algorithms (GA), Particle Swarm Optimization (PSO), Grey Wolf Optimizer (GWO), and Whale Optimization Algorithm (WOA) are successfully applied for medical data classification and optimization of neural networks [10]. Despite these advancements, application of newer, less popular bio-inspired algorithms in predicting heart disease is a scarcity.

The Aquila Optimizer (AO), driven by the hunting strategy of Aquila (eagle) birds, is a recently proposed metaheuristic with four dynamic behaviors viz [11]. exploration, exploitation, swooping, and grabbing prey to handle global and local search balance. As compared to conventional techniques, AO shows quicker convergence and increased exploration capacity, making it suitable for ANN and MLP architecture optimization. Its relative neglect in healthcare AI studies opens a new window for exploring its potential in healthcare applications.

1.1 Significance of the Study

This study situates itself at the intersection of AI, healthcare, and bio-inspired computation. It contributes to the field in the following ways:

- I. Establishing baseline predictive performance of ANN and MLP models in heart disease prediction.
- II. Investigating the application of the Aquila Optimizer as a hyperparameter tuning strategy for ANN/MLP, a novel direction that has received minimal attention in healthcare research.
- III. Comparing the tuned and untuned models using clinically relevant metrics: accuracy, precision, recall, F1-score, Matthews Correlation Coefficient (MCC), Area Under the Curve (AUC), confusion matrix, and ROC curve.
- IV. Demonstrating the viability of bio-inspired optimization in improving AI-based clinical decision support systems.

2.0 Related Works

[12] presented a model to predict the incidence of cardiac diseases with the help of machine learning techniques. They also discerned some features related to this disease. The model developed by the authors included supervised learning techniques like Naïve Bayes, decision tree, KNN, and RF. Authors utilized Cleveland dataset on cardiac disease from the UCI machine learning repository with 76 attributes and 303 instances. Only 14 of these 76 features were utilized for training and testing. After completing their experiments, authors found that KNN possessed the maximum accuracy value. [13] introduced a new approach for the quick and precise detection of abnormalities related to heart disease. Their work made use of a dataset from the UCI Machine Learning Repository, which consisted of 303 patient records and 75 features. The data was preprocessed and normalized to improve its quality and homogeneity and, subsequently, feature selection was performed to determine the most important variables for analysis. Feature extraction was applied on relevant medical images using image classification methods in a bid to uncover deeper structures. The features that were achieved were then partitioned using an Adaptive Harris Hawk Optimization (AHHO) cluster algorithm to enhance organization and interpretability of the data. The features that were clustered were used as inputs for a deep evolutionary algorithm that carried out the final classification. The proposed approach achieved 97.3% classification accuracy. Performance assessment was done on the MATLAB/Simulink platform, where the system had 95.6% accuracy, 93.8% sensitivity, and 98.6% specificity, demonstrating the efficacy and potency of the model in heart disease diagnosis. [14] presented a comprehensive methodology that includes support vector regression (SVR), neural networks, multivariate adaptive regression splines, neuro-fuzzy techniques, and statistics. They discovered that the adaptive neuro-fuzzy inference system (ANFIS) had a whopping accuracy rate of 96.56%, proving that this hybrid model excels compared to conventional models by over 90% in prediction accuracy. [15] proposed a new model for the prediction of heart disease through the integration of Genetic Algorithms (GAs) with Support Vector Machines (SVMs) to enhance precision. The research was centered on the critical area of early detection of cardiovascular diseases

through using GAs as a metaheuristic approach to optimize hyperparameters, in particular, regulating the gamma and C parameters of SVMs. Through mimicking natural selection, crossover, and mutation, the model was optimized in later generations to find best hyperparameter values. The authors noted that transforming binary chromosomes into meaningful numeric values is significant and used k-fold cross-validation to reduce error rates when evaluating the model. By applying tournament selection and mutation methods, they achieved noteworthy accuracy gains, with their performance on the heart disease dataset superior to conventional approaches, achieving an astronomical 97% accuracy in prediction.

3.0 Research Methodology

The heart disease detection modeling employed a quantitative research strategy, characterized by a systematic and structured process of constructing predictive models as shown in figure 1. It

started with pre-processing of data; wherein categorical features were one-hot encoded and numerical features normalized. This ensured data consistency, minimized bias, and improved the convergence behavior of the learning algorithm. Subsequently, the preprocessed dataset was split into individual training and test sets, thereby enabling a fair assessment of the model's predictive ability. The technical foundation was the integration of hyperparameter tuning, where the Aquila Optimizer (AO), a bio-inspired metaheuristic algorithm, was employed to exhaustively explore the hyperparameter space and identify optimal configurations for the models. The optimized hyperparameters were then employed to train a MLP and ANN algorithm on the training set. Finally, the model's predictive power and its ability to generalize were qualitatively tested on the unseen test set using standard performance metrics. This testing provided objective, quantitative proof of the model's effectiveness in heart disease prediction.

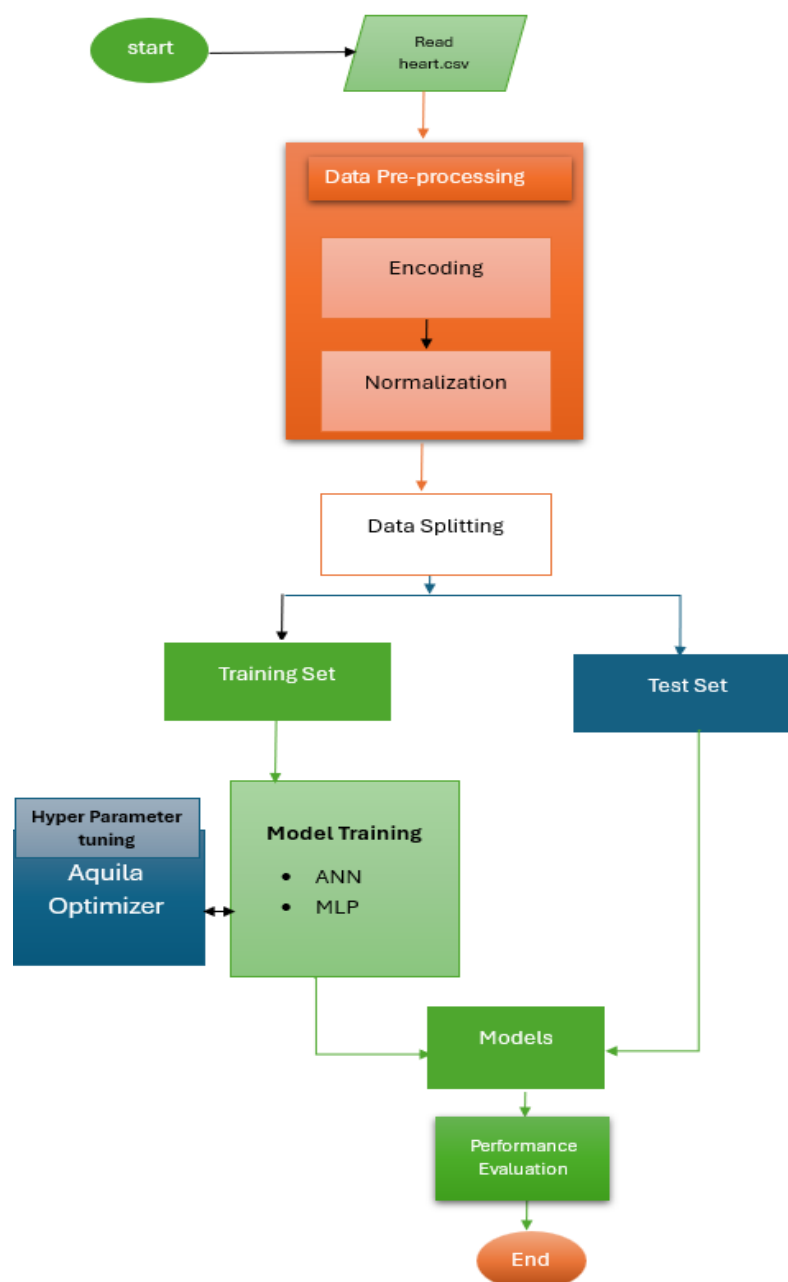


Figure 1: Research Design

3.2 Data collection and Description

The dataset employed in this research was obtained from the Kaggle Machine Learning Repository. It originates from a compilation of clinical records collected from four distinct databases: Cleveland, Hungary, Switzerland, and Long Beach V. Although the complete dataset comprises 76 attributes, most published research, including this study, focuses on a refined subset of 14 key variables considered most relevant for heart disease prediction.

3.3 Classification Models

ANN is a computational model inspired by the structure of the human brain [16]. It consists of an input layer, hidden layers, and an output layer. In the context of heart disease prediction, ANN receives patient medical features (such as age, cholesterol, blood pressure, maximum heart rate, etc.) as inputs and produces a binary output (disease = 1, no disease = 0). Mathematically, each neuron computes a weighted sum of its inputs, applies a bias, and passes the result through a nonlinear activation function. For a neuron j :

$$z_j = \sum_{i=1}^n \omega_{ij} x_i + b_j \quad (1)$$

$$a_j = f(z_j)$$

Where x_i are input features (e.g., patient attributes), ω_{ij} are trainable weights, b_j is the bias term, $f(\cdot)$ is the activation function (ReLU, Sigmoid, or Tanh), and a_j is the neuron's output.

The output layer uses the sigmoid function for binary classification:

$$\sigma(z) = \frac{1}{1 + e^{-z}} \quad (2)$$

where $\hat{y} \in [0,1]$ represents the probability of a patient having heart disease. A threshold of 0.5 is applied:

- if $\hat{y} \geq 0.5 \Rightarrow$ heart disease (Positive)
- if $\hat{y} < 0.5 \Rightarrow$ No Heart Disease (Negative)

The loss function used is the binary cross-entropy:

$$L(y, \hat{y}) = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (3)$$

where y_i is the true label (0 or 1). The optimizer (SGD, Adam, or tuned with AO) minimizes this loss to improve prediction accuracy.

The Multilayer Perceptron (MLP) is a special type of ANN consisting of multiple fully connected layers. Unlike shallow ANNs with only one hidden layer, MLP includes multiple hidden layers, allowing the network to capture more complex nonlinear relationships in patient health data [17].

For an MLP with L layers, the forward propagation is given as:

$$z^{[l]} = W^{[l]} a^{[l-1]} + b^{[l]}, \quad l = 1, 2, \dots, L \quad (4)$$

$$a^{[l]} = f(z^{[l]})$$

where: $W^{[l]}$ and $b^{[l]}$ are weights and biases of layer l , $a^{[l-1]}$ is the activation from the previous layer, $f(\cdot)$ is an activation function (ReLU in hidden layers, Sigmoid in the final layer). The output layer is identical to ANN's binary classification case:

$$\hat{y} = \sigma(W^{[L]} a^{[L-1]} + b^{[L]}) \quad (5)$$

3.4 Model Hyper-Parameter Tunning

The Aquila Optimizer (AO) is a recent bio-inspired optimization algorithm proposed in 2021 as a strategy that emulates the clever hunting strategy of Aquila eagles in nature [18]. Its novelty lies in its ability to balance the exploration and exploitation phases, which are two vital phases in addressing global optimization problems [19]. Exploration enables the algorithm to thoroughly explore the solution space trying to avoid local minima, while exploitation attempts to enhance solutions nearby areas of interest. This balance makes AO especially suitable for addressing high-dimensional, non-convex optimization problems such as deep learning hyperparameter tuning and feature selection in medical diagnostics.

The algorithm models four distinct hunting strategies of Aquila eagles as shown in algorithm 1:

- (i) high soar with vertical stoop (exploration)
 - (ii) (contour flight with short glide attack (exploitation),
 - (iii) low flight with slow descent attack (transition), and
 - (iv) swooping with grabbing prey (intensification).
- These strategies are mathematically encoded into update rules that iteratively refine candidate solutions over a defined search horizon.

Algorithm 1: Aquila Optimizer (AO) for MLP & ANN Hy-parameter tuning**Input:**

N = Population size
 T = Maximum iterations
 d = Dimension of solution space (number of hyperparameters/features)
 LB, UB = Lower and upper bounds of each hyperparameter
 λ = Weight balancing accuracy and feature reduction
 Fitness function $F(S) = \lambda \cdot (1 - Acc(S)) + (1 - \lambda) \cdot \frac{|S|}{d}$

Output:

Best hyperparameters (x_{best}) and corresponding fitness value

1. Initialize population $X = \{X_1, X_2, \dots, X_N\}$ randomly within bounds $[LB, UB]$
2. Evaluate fitness of each X_i using ANN/MLP performance
3. Set x_{best} = best solution with minimum fitness
4. For iteration $t = 1$ to T do
 - Compute mean position x_{mean} of all solutions
 - For each candidate X_i in population:
 - Generate random numbers $r, \delta, \alpha, \beta, A$
 - Choose a hunting strategy based on probability:
 - a) Exploration (High soar & vertical stoop):

$$X^{t+1} = X_{best}^t \cdot (1 - \frac{t}{T}) + (X_{mean}^t - X_i^t) \cdot r$$
 - b) Exploitation (Contour flight & glide attack):

$$x_{rand} = \text{random solution from population}$$

$$X^{t+1} = X_{best}^t + (X_{rand}^t - X_i^t) \cdot A$$
 - c) Transition (Low flight & slow descent):

$$X^{t+1} = (X_{best}^t - X_{mean}^t) \cdot \alpha - \beta \cdot X_i^t$$
 - d) Intensification (Swooping & grabbing prey):

$$X^{t+1} = X_{best}^t + \delta \cdot (UB - LB) \cdot r$$
 - Ensure X^{t+1} respects boundary constraints $[LB, UB]$
 - End For
 - Evaluate new population fitness
 - Update x_{best} if better solution found
 - Record best fitness value for iteration t
- End For

5. Return x_{best} and corresponding ANN/MLP accuracy and metrics

Mathematically, the AO begins by initializing a population of candidate solutions within defined search space boundaries. If the optimization problem is defined in a d – dimensional space, each candidate solution can be expressed as a vector:

$$X_i = (x_{i1}, x_{i2}, \dots, x_{idd}), i = 1, 2, \dots, N \quad (6)$$

where N is the population size and d is the number of decision variables. The optimization process is guided by updating these candidate solutions iteratively over a maximum number of iterations T . At each iteration t , the best solution so far, denoted as x_{best}^t , is retained to influence the movement of other agents.

The exploration phase, corresponding to the Aquila's high soar and vertical stoop, is formulated as:

$$X^{t+1} = X_{best}^t \cdot (1 - \frac{t}{T}) + (X_{mean}^t - X_i^t) \cdot r \quad (7)$$

where X_{mean}^t is the mean position of the population at iteration t , r is a uniformly distributed random variable in $[0,1]$, and $(1 - \frac{t}{T})$ ensures a gradual reduction in exploration over time. This phase allows the algorithm to search new areas of the solution space broadly.

The exploitation phase, representing contour flight and glide attack, is described as:

$$X^{t+1} = X_{best}^t + (X_{rand}^t - X_i^t) \cdot A \quad (8)$$

where X_{rand}^t is a randomly selected solution from the population and A is a flight adjustment factor that reduces the step size over time. This mechanism focuses on refining solutions near promising regions of the search space.

The transition phase, simulating low flight and slow descent attack, is expressed as:

$$X^{t+1} = (X_{best}^t - X_{mean}^t) \cdot \alpha - \beta \cdot X_i^t \quad (9)$$

where α and β are control coefficients that balance exploration and exploitation. This phase ensures a smooth shift from exploration to exploitation, preventing premature convergence.

Finally, the swooping and grabbing prey strategy is modeled as:

$$X^{t+1} = X_{best}^t + \delta \cdot (UB - LB) \cdot r \quad (10)$$

where UB and LB are the upper and lower bounds of the search space, δ is a random perturbation factor, and r ensures stochastic variation. This intensifies the search around the best solution found so far, allowing AO to fine-tune solutions effectively.

In this research, AO was integrated for parameter tuning of Artificial Neural Network (ANN) and Multi-Layer Perceptron (MLP) models. Each candidate solution encoded both a set of hyperparameters (e.g., learning rate, number of hidden layers, neurons per layer, activation functions, and batch size).

The optimization process was guided by a fitness function that simultaneously maximized classification performance and minimized model complexity. The fitness function was expressed as:

$$F(S) = \lambda \cdot (1 - Acc(S)) + (1 - \lambda) \cdot \frac{|S|}{d} \quad (11)$$

where $Acc(S)$ is the classification accuracy obtained using the selected feature subset S , $|S|$ is the number of features in the subset, d is the total number of available features, and λ in $[0,1]$ is a weighting factor that balances accuracy maximization against dimensionality reduction. This formulation ensures that AO searches for compact yet highly discriminative feature subsets.

With AO for parameter optimization, the tuned ANN and MLP models were then evaluated on the heart disease dataset using a comprehensive suite of metrics: accuracy, precision, recall, F1-score, Matthews Correlation Coefficient (MCC), AUC, ROC curve, and confusion matrix. Comparative experiments were conducted between baseline models (without AO) and tuned models (with AO).

3.4.1 Performance Evaluation

To evaluate the performance of the ANN and MLP models. The metric presented in table 1 where used base on the definition of the TN, TP, FP and FN variables where:

- i. TP = True Positives (correctly predicted cardio cases)
- ii. TN = True Negatives (correctly predicted non-cardio cases)
- iii. FP = False Positives (non-cardio cases misclassified as cardio)
- iv. FN = False Negatives (cardio cases misclassified as non-cardio)

Metrics	Formulae	Interpretation
Accuracy Score	$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$	Define the proportion of correctly classified samples among all sample
Precision	$precision = \frac{TP}{TP + FP}$	Of all predicted cardio cases, precision define how many were correct.
Recall	$Recall = \frac{TP}{TP + FN}$	Ability of the model to correctly identify cardio patients.
F1-Score	$F - Measure = 2 \times \frac{Precision \times Recall}{Precision + Recall}$	Defines the harmonic mean of precision and recall, balancing false positives and false negatives.
Area under the Curve	$AUC = \frac{1}{2} \left(1 + \frac{TP}{TP + FN} + \frac{FP}{FP + FN} \right)$	Scalar measure of separability. AUC = 1.0 means perfect classification, while AUC = 0.5 corresponds to random guessing.
Matthews Correlation Coefficient (MCC)	$MCC = \frac{(TP.TN) - (FP.FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$	Balanced metric even when classes are imbalanced. Values range from -1 (total disagreement) to +1 (perfect prediction), with 0 meaning random guessing.

4.1 Environmental Setup

The predictive model was developed using the Anaconda computing platform. Testing was conducted on Windows operating system with dual-core Intel Core i5 processor and 8 GB RAM. Deployment was conducted in Python using the standard machine learning libraries Scikit-Learn, TensorFlow, NumPy, and pandas.

4.2 Parameter Settings

Table 2: Parameter Settings

Parameter	Value
Population Size	8
Max Iterations	10
Learning Rate Bounds	[0.0001, 0.01]
Units Bounds	[32, 128]
Dropout Bounds	[0.1, 0.5]
Batch Size Bounds	[16, 64]
Parameter	Value

The population size 8 is the number of candidate solutions (individuals) maintained in each step of the AO search process. A small population size was utilized to balance computational efficiency and exploration capability, especially since deep learning optimization may prove to be computationally expensive. The number of maximum iterations (10) specifies the number of search steps for the optimizer. While larger iteration sizes may reach higher convergence, preliminary testing indicated that ten iterations were sufficient to produce stable performance improvement within the computational budget. The limits of the learning rate ([0.0001, 0.01]) allow the optimizer to look for anything in between very small steps (enhancing stability and high-resolution learning) and larger strides (raising the convergence speed). It is widely referred to throughout deep learning literature as an appropriate balance between speed and training stability. The limits for units ([32, 128]) determine the sizes of the hidden layers in the ANN/MLP. Small networks risk underfitting the data, while extremely large networks increase the likelihood of overfitting and computational cost. By allowing AO to search within this bounded space, the model achieves balance between generalization and expressiveness.

In this study, the AO was employed to tune the hyperparameters of the ANN and MLP models for heart disease prediction. The performance of bio-inspired optimizers is highly dependent on the initialization of their search space and parameter ranges. To ensure a fair and controlled optimization process, the parameters of AO were carefully defined as shown in Table 2.

The dropout thresholds ([0.1, 0.5]) make regularization flexible. Dropout prevents overfitting by randomly shutting off a percentage of the neurons during training. Lower than 0.1 (near 0.1) preserves most of the network structure, while higher values (max up to 0.5) impose more regularization. AO selected dropout values independently based on their effect on fitness, hence achieving balance between bias and variance.

Lastly, the batch size boundaries ([16, 64]) dictate the number of samples that are passed through before a weight update. Smaller batch sizes generalize more by introducing noise to gradient estimation, while larger batch sizes stabilize the updates and utilize computational resources better. The AO configuration provides the ability to explore throughout this interval for identifying the batch size that is best at obtaining model performance on the validation set.

4.3 Result Performance Evaluation

The experimental results from table 3 and figure 2 provide a comprehensive comparison between baseline models (MLP and ANN) and their Aquila Optimizer (AO)-tuned counterparts. The metrics considered include the accuracy, precision, recall, f1-score, MCC, and AUC.

Table 3: Models Performance Evaluation

Model	Accuracy	Precision	Recall	F1-score	MCC	AUC
Baseline MLP	0.8537	0.8288	0.8932	0.8598	0.7094	0.8535
Baseline ANN	0.9561	0.9434	0.9709	0.9569	0.9126	0.956
AO-Tuned MLP	0.9805	1.0	0.9612	0.9802	0.9617	0.9806
AO-Tuned ANN	0.9854	1.0	0.9709	0.9852	0.9712	0.9854

Considering the baseline models, the baseline MLP achieved an accuracy of 0.8537, precision of 0.8288, recall of 0.8932, F1-score of 0.8598, Matthews Correlation Coefficient (MCC) of 0.7094, and an AUC of 0.8535. These results indicate that while the MLP was fairly effective at classifying heart disease patients, there were some misclassifications, particularly in precision, which shows that the model had some false positives. The baseline ANN significantly outperformed the MLP across all metrics. It attained an accuracy of 0.9561, precision of 0.9434, recall of 0.9709, F1-score of 0.9569, MCC of 0.9126, and an AUC of 0.9560. The high recall demonstrates that the ANN was highly effective at correctly identifying patients with heart disease, while the high MCC and F1-score suggest a strong balance between positive and negative predictions

After applying the Aquila Optimizer for hyperparameter tuning with a stable fitness value around 0.0549, both models experienced remarkable performance improvements. The AO-tuned MLP improved substantially with an accuracy of 0.9805, perfect precision of 1.0000, recall of 0.9612, F1-score of 0.9802, MCC of 0.9617, and an AUC of 0.9806. The perfect precision indicates that

the tuned MLP had no false positives, which is particularly valuable in medical diagnostics, as it reduces the likelihood of misclassifying healthy individuals as diseased. The AO-tuned ANN achieved the best overall performance with an accuracy of 0.9854, precision of 1.0000, recall of 0.9709, F1-score of 0.9852, MCC of 0.9712, and AUC of 0.9854. Compared to the baseline ANN, this represents a significant improvement in all metrics. The combination of perfect precision, very high recall, and excellent MCC indicates that the AO-tuned ANN is both robust and generalizable, making it a highly reliable model for heart disease prediction.

Conclusively, it can be deduced that the ANNs inherently outperform MLPs in this heart disease classification task, even before tuning. Furthermore, AO significantly enhances performance by finding better hyperparameter configurations, especially evident in the precision (1.0000) and MCC (>0.96) values. Hence, the AO-tuned ANN is the most reliable model, combining high accuracy, zero false positives, and strong generalization metrics.

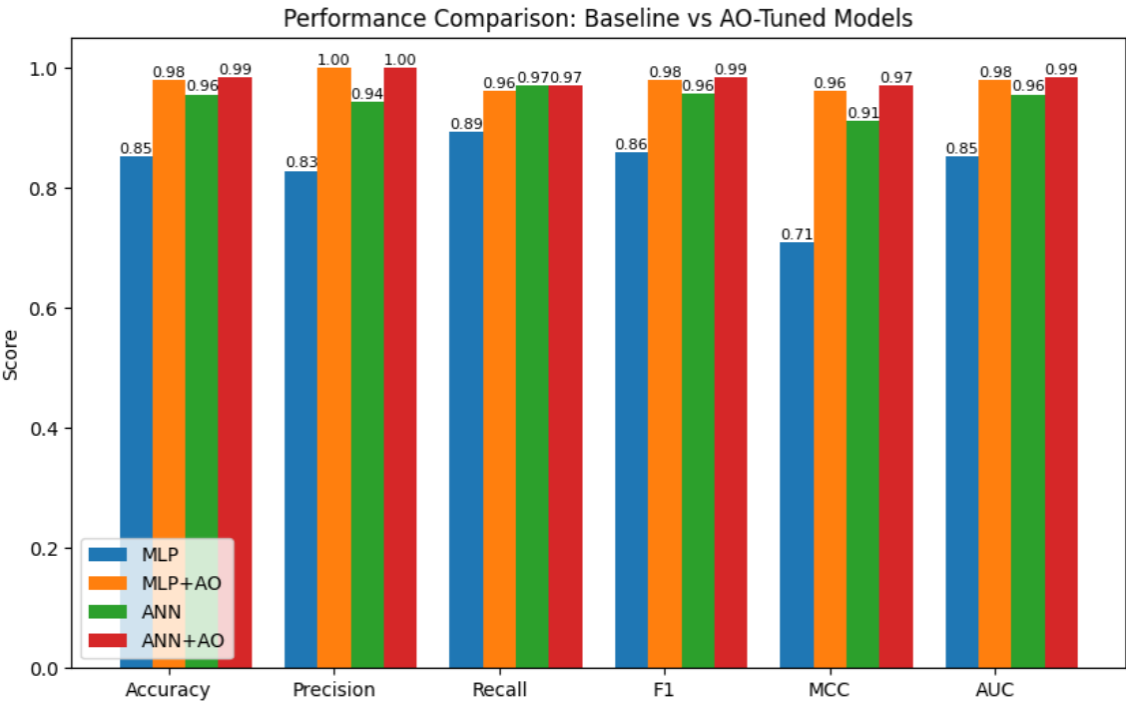


Figure 2: Result Comparison

5.0 Conclusion

This research has examined the application of ANNs and MLPs in the prediction of heart disease, where the need to fine-tune hyperparameters is vital to maximizing diagnostic performance. The results confirm that ANNs improve naturally over MLPs due to their capacity for representing higher-order non-linear interactions between patient variables. However, the sensitivity of both algorithms to hyperparameter settings highlights the need for efficient optimization strategies. With the addition of the AO, a newly explored and less-studied bio-inspired algorithm, this research was able to demonstrate outstanding performance improvements across all measures of classification. The models tuned using AO outperformed their baselines, with the tuned ANN demonstrating near-perfect classification performance with 98.54% accuracy, 100% precision, and MCC of 0.9712. The tuned MLP also demonstrated significant improvements, with 100% precision, eliminating false positives, and becoming even more dependable in clinical practice. Lastly, the outcome confirms AO to be an efficient hyperparameter tuner which will be superior to the shortcomings of the traditional search algorithms such as grid and random search. In practical terms, these results portend the potential of AO-tuned neural networks as extremely useful decision aids in health care. With strong predictive accuracy, no false positives, and respectable generalization, the AO-tuned ANN is especially well-suited for clinical application, where accuracy in diagnosis has direct patient outcome implications.

Essentially, this research contributes to revealing the Aquila Optimizer as a novel, high-performance hyperparameter optimization method in medical AI systems. This contribution paves the way for the further exploration of bio-inspired optimization in healthcare, which could extend to high-volume datasets, multimodal data fusion, and real-time clinical decision-making platforms.

References

1. Muhu, B. W. (2024). Risk Factors Associated with Cardiovascular Diseases Among Patients Attending Nyeri County Referral Hospital, Central Kenya.
2. Inbar, O., Inbar, O., Dlin, R., & Casaburi, R. (2025). Transitioning from stress electrocardiogram to cardiopulmonary exercise testing: a paradigm shifts toward comprehensive medical evaluation of exercise function. *European Journal of Applied Physiology*, 1-12.
3. Eskandar, K. (2023). Artificial intelligence in healthcare: explore the applications of AI in various medical domains, such as medical imaging, diagnosis, drug discovery, and patient care. *Series Med Sci*, 4, 37-53.
4. Rimal, Y., Sharma, N., Paudel, S., Alsadoon, A., Koirala, M. P., & Gill, S. (2025). Comparative analysis of heart disease prediction using logistic regression, SVM, KNN, and random forest with cross-validation for improved accuracy. *Scientific Reports*, 15(1), 13444.
5. Alsabhan, W., & Alfadhly, A. (2025). Effectiveness of machine learning models in diagnosis of heart disease: a comparative study. *Scientific Reports*, 15(1), 24568.
6. Maji, K., Gupta, S., & Dutta, P. K. (2024, December). Enhancing heart disease prediction accuracy: comprehensive analysis of XGBoost and AdaBoost. In *IET Conference Proceedings CP913* (Vol. 2024, No. 37, pp. 130-136). Stevenage, UK: The Institution of Engineering and Technology.
7. Ahmed, S. F., Alam, M. S. B., Hassan, M., Rozbu, M. R., Ishtiak, T., Rafa, N., ... & Gandomi, A. H. (2023). Deep learning modelling techniques: current progress, applications, advantages, and challenges. *Artificial Intelligence Review*, 56(11), 13521-13617.
8. Kadhim, Z. S., Abdullah, H. S., & Ghathwan, K. I. (2022). Artificial Neural Network Hyperparameters Optimization: A Survey. *Int. J. Online Biomed. Eng.*, 18(15), 59-87.
9. Ibrahim, M. Q., Hussein, N. K., Guinovart, D., & Qaraad, M. (2025). Optimizing Convolutional Neural Networks: A Comprehensive Review of Hyperparameter Tuning Through Metaheuristic Algorithms. *Archives of Computational Methods in Engineering*, 1-38.
10. Kumar, N., & Kumar, D. (2021). An improved grey wolf optimization-based learning of artificial neural network for medical data classification. *Journal of Information and Communication Technology*, 20(02), 213-248.
11. Sharma, B., Rizwan, M., & Anand, P. (2023). Optimal design of renewable energy-based hybrid system considering weather forecasting using machine learning techniques. *Electrical Engineering*, 105(6), 4229-4249.
12. Shah, D., Patel S., & Bharti S.K. (2020). Heart Disease Prediction using Machine Learning Techniques. *SN Comput. Sci.* 2020; 1:1-6.
13. Balamurugan, R., Ratheesh S. & Venila Y.M. (2022). Classification of heart disease using adaptive Harris hawk optimization-based clustering algorithm and enhanced deep genetic algorithm. *Soft Comput.* 2022;(26)2357-2373.
14. Taylan, O., Alkabaa, A. S., Alqabbaa, H. S., Pamukcu, E., & Leiva, V. (2023). Early prediction in classification of cardiovascular diseases with machine learning, neuro-fuzzy and statistical methods. *Biology*, 12(1), 117.
15. PALIWAL, S., PRAVEEN, S., ALAM, M. A., & AHMED, J. (2024). Optimizing heart disease prediction models using genetic algorithms: a metaheuristic approach. *Journal of Theoretical and Applied Information Technology*, 102(9).
16. Qamar, R., & Zardari, B. A. (2023). Artificial neural networks: An overview. *Mesopotamian Journal of Computer Science*, 2023, 124-133.
17. Desai, M., & Shah, M. (2021). An anatomization on breast cancer detection and diagnosis employing multi-layer perceptron neural network (MLP) and Convolutional neural network (CNN). *Clinical eHealth*, 4, 1-11.
18. Cui, H., Xiao, Y., Hussien, A. G., & Guo, Y. (2024). Multi-strategy boosted Aquila optimizer for function optimization and engineering design problems. *Cluster Computing*, 27(6), 7147-7198.
19. Sasmal, B., Hussien, A. G., Das, A., & Dhal, K. G. (2023). A comprehensive survey on aquila optimizer. *Archives of Computational Methods in Engineering*, 30(7), 4449-4476.