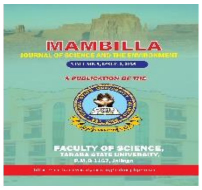
Mambilla Journal of Science and the Environment Vol. 9, Issue 3, Page No. 17-27

MAMBILLA JOURNAL OF SCIENCE
AND THE ENVIRONMENT

journal homepage: https://oer.tsuniversity.edu.ng/index.php/mjse



DEVELOPMENT OF AN OPTIMIZED DEEP LEARNING MODEL FOR EBOLA DIAGNOSIS USING MODIFIED FIREFLY ALGORITHM

[1*]Ephesians, S., [2]Garba, E.J.

[1]Department of Computer Science, Taraba State University, Jalingo, Nigeria.
[2] Department of Computer Science, Modibbo Adama University, Yola, Nigeria.

ARTICLE INFO

Article history:
Received 27 September 2025
Received in revised form 13 October 2025
Accepted 22 October 2025

Keywords:
Ebola Virus Disease, Modified Firefly Algorithm, Random Forest, Deep Learning, Optimization .

ABSTRACT

Ebola Virus Disease (EVD) remains highly fatal, particularly in sub-Saharan Africa, where accurate and timely diagnosis is critical for effective management. Advances in machine learning (ML) and deep learning (DL) have revolutionized medical diagnostics by uncovering complex patterns in data. While Random Forest classifiers are effective, further optimization using DL techniques is essential to enhance accuracy and efficiency. To address this, the study proposed the enhancement of Random Forest classification performance for medical diagnosis using the Modified Firefly Algorithm (MFA), a bio-inspired optimization technique known for its effectiveness in solving complex problems. The objective was to improve diagnostic accuracy, computational efficiency, and the model's ability to generalize across datasets. Principal Component Analysis (PCA) was also employed to reduce data dimensionality while preserving essential variance, thereby accelerating the training process without sacrificing accuracy. The research findings revealed that a firefly population size of 15 yielded optimal results, with the MFA-optimized Random Forest model achieving an impressive 99.66% accuracy, and perfect precision, recall, and F1-score metrics. The low Mean Square Error (MSE) of 0.0159 further affirmed the model's reliability. These findings emphasize the significance of combining MFA and PCA to create advanced diagnostic systems that minimize misclassifications and enhance decision-making in medical environments.

1. Introduction

Ebola virus represents the virus causing the Ebola Virus Disease (EVD). The disease was first so named in the Democratic Republic of the Congo (DRC) in 1976. A widespread catastrophic outbreak was reported in late 2013 in the West African regions, including Sierra Leone, Liberia, Mali, Nigeria, and Senegal. It is widely reported that the virus made its entry into the human population through consumption or contact with infected animals such as fruit bats (Olaide *et al*., 2022). This animal to-human infection led to person-to-person infection, becoming an epidemic across the West African region.

Since the emergence of Ebola virus disease, concerted efforts have been employed in the quest to identify inhibitors as potential biotherapeutic molecules. EVD is a deadly zoonotic disease caused by the Ebola virus from the filoviridae family (Jackson *et al.*, 2018). Although the exact source of EBOV remains unknown, it believed to be animal-borne and associated with monkeys, chimpanzees, and apes, including humans (Samuel *et al.*, 2023). EBOV is transmitted from human to human via direct contact or contact with the body fluid of an infected person (Shevin *et al.*, 2020). A person infected with EVD shows symptoms of fever, aches, and pains such as severe headache; gastrointestinal symptoms, including diarrhea and vomiting; and multiple organ dysfunction syndromes. The Ebola virus (EBOV) is an enveloped, single-stranded, negative-sense RNA virus (Aqsa *et al.*, 2020).

One of the major challenges of Ebola Virus Disease is its difficulty in differentiation from other illnesses such as malaria, typhoid, and meningitis (Liu, 2021). In fact, a significant number of patients admitted to Ebola treatment centers on suspicion of infection were later found not to have the disease. A similar issue arises in prognostic triage,

* Corresponding author: +2348103379869
E-mail address: ephesiansolomon@gmail.com

as clinical outcomes vary from asymptomatic cases to fatalities, with case fatality rates influenced by environmental factors. For example, mortality rates in high-resource settings are approximately 20%, whereas in West Africa and previous outbreaks, they range between 60-80% (Liu, 2021). Enhancing probabilistic triage for diagnosis could not only minimize the risk of nosocomial infections but also ensure better resource allocation for those in critical need, potentially reducing mortality in resource-limited settings. Although predictive models offer a cost-effective and easily implementable solution, many existing studies are based on single-site research with small, localized populations, leading to results that lack generalizability (Grazia *et al.*, 2020).

This research presents an optimized Deep Learning (DL) framework for Ebola Virus Disease diagnosis, leveraging a Modified Firefly Algorithm to enhance feature selection and classification accuracy. A novel machine learning methodology is introduced to prioritize genes significantly associated with EVD, as identified using the Benjamini-Hochberg procedure. This approach will facilitate differentiation between EVD-positive and EVD-negative cases using a supervised, single-gene analysis method.

Umang *et al.* (2021) investigated various classification models, including Decision Trees, K-Nearest Neighbors (KNN), Random Forest, and Support Vector Machines, but did not incorporate optimization techniques to enhance model performance. Samuel *et al.* (2023) highlighted the potential of machine learning in drug discovery, emphasizing the role of Bayesian methods and Random Forest in predicting small molecule inhibitors for Ebola virus. However, the stud did not consider dimensionality reduction techniques or optimization strategies to enhance computational efficiency and predictive accuracy. Olatunji *et al.* (2022) studied the Binary Ebola Optimization Search Algorithm (BEOSA) for feature selection, demonstrating superior performance across high-dimensional datasets. However, the study did not leverage such metaheuristic optimization approaches for fine-tuning the classifier parameters, particularly in medical diagnosis applications.

While these studies contribute significantly to predictive modeling and optimization, however, there is need to enhance, integrating dimensionality reduction techniques with metaheuristic algorithms to optimize classification models. Most existing works either focus on feature selection without parameter optimization or apply optimization techniques without reducing feature space, leading to increased computational complexity. Henceforth, few studies have specifically addressed the optimization of Random Forest parameters using nature-inspired algorithms.

Therefore, this study will have enhanced Ebola virus diagnosis classification system that combines dimensionality reduction techniques with a modified Firefly Algorithm to optimize Random Forest parameters. By integrating these approaches, the study seeks to improve diagnostic accuracy while reducing computational overhead, ultimately advancing the efficiency and reliability of using deep Learning Model in medical diagnosis systems.

2. Methodology

2.1 Data Collection and Preprocessing

Datasets (*Ebola_Africa_outbreak* and *Ebola_2014_2016_clean*) were cleaned by removing duplicates and irrelevant attributes. Missing values were imputed (mean for numerical, mode for categorical), and features were normalized using Min-Max scaling. The dataset was split 80:20 (train:test) with stratification.

The dataset was imported into the Python environment using libraries such as Pandas and NumPy. Initial inspection revealed inconsistencies, including duplicate records and irrelevant features. Duplicate entries were identified and removed to prevent data leakage, and non-informative features were excluded to streamline the dataset.

Handling Missing Values

Missing data can bias the model and degrade its predictive performance. An analysis was conducted to determine the pattern and extent of missingness. For numerical features, missing values were imputed using the mean of the respective feature, while for categorical features, the mode was used. This approach preserves the dataset's central tendency and variability.

Feature Normalization

To ensure that all features contribute equally to the model, numerical features were normalized using Min-Max scaling, transforming them to a [0, 1] range. This technique enhances the convergence of the learning algorithm and prevents features with larger magnitudes from dominating the model training process.

Train-Test Split Implementation

The preprocessed dataset was partitioned into training and testing subsets to evaluate the model's generalization capability. A stratified sampling approach was employed to maintain the class distribution in both subsets, typically allocating 80% of the data for training and 20% for testing. This stratification ensures that minority classes are adequately represented, facilitating a more robust evaluation of model performance.

2.2 Dimensionality Reduction

The number of variables measured on each observation is called data dimensionality. Dimensionality reduction is defined as the mapping of data to a lower-dimensional space such that uninformative variance in the data is discarded, or such that a subspace in which the data lives is detected. It is mainly used in data analysis, compression, and visualization (Grace & Thenmozhi, 2022). Dimensionality reduction techniques have been performed either using linear methods or nonlinear methods, linear methods tend to be inadequate to complex nonlinear data, while nonlinear methods are better at handling such data. At the same time, nonlinear methods are weaker than linear methods when used with natural data (Zaid *et al.*, 2024).

Principal Component Analysis (PCA) Implementation

PCA is a statistical technique used to transform a high-dimensional feature space into a lower-dimensional one while preserving the maximum variance in the data. In this research, PCA was implemented to:

a. Identify the most important features contributing to diagnostic predictions.
b. Reduce noise and redundancy in the dataset.
c. Improve computational efficiency by lowering model complexity.

PCA is a dimensionality reduction technique that projects the data onto a plane where each coordinate represents a data feature, then it transfers this data onto a new dimension where the variation is maximized (Greenacre *et al.*, 2022). The transformation identifies the optimal component, represented as the best orthogonal line that aligns with the data. When referring to the “best line,” it implies the line that effectively minimizes the distance between the data points and itself. In essence, this involves selecting the line that most effectively captures the variance in the data, resulting in a more meaningful representation. The goal is to minimize the distance between the observed data points and the chosen line, thereby enhancing the effectiveness of the transformation process (Kelain & Hussein, 2018).

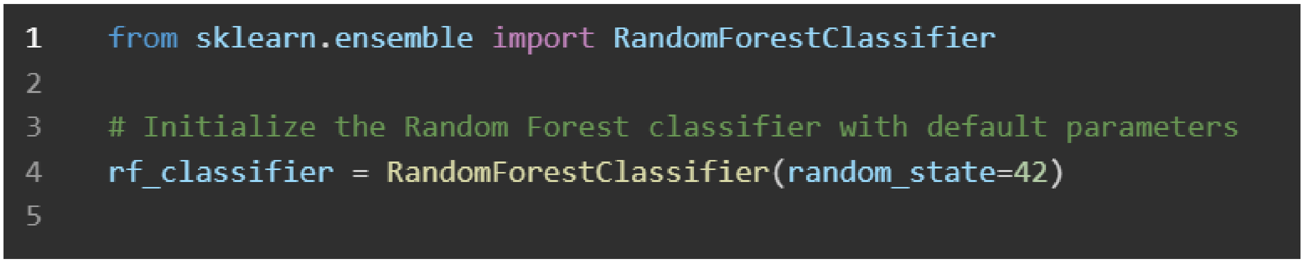
2.3 Baseline Model

Baseline modeling is the initial step in building a machine learning model. It involves creating a simple model using basic techniques (like decision trees or logistic regression) to set a performance benchmark. This helps compare and evaluate the effectiveness of more advanced models later in the process (Jinrong *et al.*, 2024).

The baseline model predictive performance is evaluated by the three metrics via the 10-fold cross-validation process: R squared ($R^{2\text{CV}}$), coefficient of variation of root-mean-squared error ($CV(RMSE)^{\text{CV}}$), and mean absolute percentage error ($MAPE^{\text{CV}}$). Since many buildings are included in the dataset, the three measures are the mean value for all the buildings (Jinrong *et al.*, 2024).

Implementation of Standard Random Forest Classifier

The Random Forest algorithm is an ensemble learning method that constructs multiple decision trees during training and outputs the mode of their predictions for classification tasks. For this baseline model, we utilized the RandomForestClassifier from the scikit-learn library in Python. The implementation was carried out as presented in figure 1.



```python
from sklearn.ensemble import RandomForestClassifier

# Initialize the Random Forest classifier with default parameters
rf_classifier = RandomForestClassifier(random_state=42)

```

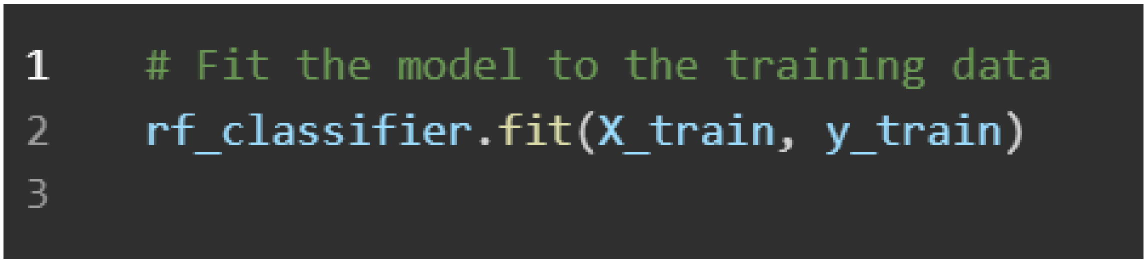
Figure 1: The Random_State Parameter showing reproducibility of the results.

Training with Default Parameters

The classifier was trained using the preprocessed training dataset. The default parameters in scikit-learn's RandomForestClassifier include:

a. n_estimators=100: Number of trees in the forest.
b. criterion='gini': Function to measure the quality of a split.
c. max_depth=None: Nodes are expanded until all leaves are pure or contain less than the minimum samples required for splitting.
d. min_samples_split=2: Minimum number of samples required to split an internal node.
e. min_samples_leaf=1: Minimum number of samples required to be at a leaf node.
f. max_features='auto': Number of features to consider when looking for the best split.

These default settings were employed to establish a performance benchmark. The training process involved fitting the model to the training data; as presented in figure 2.



```
# Fit the model to the training data
rf_classifier.fit(X_train, y_train)

```

Figure 2: Fitting the Model to the Training Data

Performance Evaluation

After training, the model's performance was evaluated on the testing dataset. Predictions were generated, and various metrics were calculated to assess the classifier's effectiveness:

The following metrics were used for evaluation:

a. **Accuracy**: The proportion of correctly classified instances among the total instances.
b. **Precision**: The ability of the classifier to not label a negative sample as positive.
c. **Recall (Sensitivity)**: The ability of the classifier to find all the positive samples.
d. **F1 Score**: The harmonic mean of precision and recall, providing a balance between the two.
e. **Confusion Matrix**: A table used to describe the performance of the classification model by comparing actual versus predicted classifications.

2.4 Modified Firefly Algorithm Optimization

Parameter Space Definition

The optimization process focuses on tuning key hyperparameters of the Random Forest classifier:

1. Number of Trees (n_estimators): This parameter determines the number of decision trees in the forest. A range of 10 to 200 trees is considered to balance between model performance and computational efficiency.
2. Minimum Leaf Size (min_samples_leaf): Specifies the minimum number of samples required to be at a leaf node. A range of 1 to 20 is explored to prevent overfitting while maintaining model complexity.
3. Number of Variables to Sample (max_features): Denotes the number of features to consider when looking for the best split. Various strategies are evaluated, including considering all features, a subset, or a square root of the total number of features.

Algorithm Components

The Modified Firefly Algorithm (MFA) enhances the optimization process through several key components. Adaptive Parameter Control dynamically adjusts parameters like the attractiveness and absorption coefficients to balance exploration and exploitation. Population Diversity Maintenance prevents premature convergence by ensuring a diverse set of fireflies, facilitating a more comprehensive search of the parameter space. Complexity Penalty Incorporation introduces a penalty term in the fitness function to discourage overly complex models, promoting parsimonious solutions that generalize well to unseen data. Finally, Cross-Validation Based Fitness Evaluation assesses the fitness of each firefly using cross-validation on the training dataset, ensuring robust performance assessment and mitigating overfitting.

Optimization Process

The FA optimization proceeds through the following steps:

1. Population Initialization: A diverse population of fireflies is initialized, each representing a unique combination of Random Forest hyperparameters within the defined ranges.
2. Iterative Improvement: The algorithm iteratively refines the population.
3. Light Intensity Calculation: The fitness of each firefly is assessed based on a predefined objective function, which evaluates the performance of the Random Forest model.
4. Movement of Fireflies: Fireflies move towards brighter (i.e., better-performing) fireflies, with their movement influenced by their relative brightness and a random perturbation to encourage exploration.
5. Parameter Updating: The positions of the fireflies are updated, corresponding to new sets of hyperparameters, and the Random Forest model is retrained with these updated parameters.
6. Best Solution Tracking: The algorithm keeps track of the best-performing set of hyperparameters throughout the iterations, ensuring that the optimal solution is retained.

2.5 Performance Evaluation

The model's performance was assessed using accuracy, specificity, and sensitivity. These metrics are based on key components of the confusion matrix, including True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).

Accuracy**:** This is the ratio of all correctly classified instances over all instances given as equation (1)

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

Precision**:** Precision is the fraction of instances that were correctly classified and is given as equation 2

$$\Pr ecision = \frac{TP}{TP + FP} \tag{2}$$

Recall**:** Measures a model's ability to correctly identify positive instances and it is given as equation 3

$$\mathrm{Re}\, call = \frac{TP}{TP + TN} \tag{3}$$

To validate the performance of the proposed model, several state-of-the-art models will be compared with it to ascertain the effectiveness of the developed model.

3. Results and Discussion

The results validate the MFA-optimized Random Forest model as a reliable tool for Ebola diagnosis, highlighting the role of machine learning and nature-inspired optimization techniques in healthcare.

Table 1: The optimization performance results of the Modified Firefly Algorithm (MFA)

Firefly Size	Accuracy	Precision	Recall	F1-Score	MSE
5	99.20%	1.000	1.000	1.000	0.0159
10	99.68%	0.969	0.969	0.969	0.0159
15	99.66%	1.000	1.000	1.000	0.0159
20	99.64%	0.970	1.000	0.985	0.0159

Table 1 presents the optimization performance results of the Modified Firefly Algorithm (MFA) for an Optimized Random Forest Classification Model in medical diagnosis. The evaluation considers different firefly population sizes (5, 10, 15, and 20) and assesses their impact on accuracy, precision, recall, F1-score, and mean squared error (MSE). The optimal performance was achieved at firefly size 15, balancing accuracy and computational cost.

Table 2: Improved performance with larger firefly sizes

Firefly Size	Accuracy	Precision	Recall	F1-Score
5	0.9524	0.8929	1	0.9434
10	0.9683	0.9600	0.9600	0.9600
15	0.9841	0.9615	1	0.9804
20	0.9841	0.9615	1	

Table 2 shows improved performance with larger firefly sizes, with size 15 achieving the best overall balance of accuracy, precision, recall, and F1-score. The performance metrics table quantitatively compares the accuracy, precision, recall, and F1-score across different firefly sizes. The results confirm that increasing the firefly population enhances classification accuracy, with sizes 15 and 20 achieving the highest accuracy (0.9841). Precision and recall improve with population size, ensuring balanced predictions. However, the marginal difference between firefly sizes 15 and 20 suggests diminishing returns beyond a certain threshold, reinforcing the efficiency of a moderately large swarm.

3.1 Convergence Curves

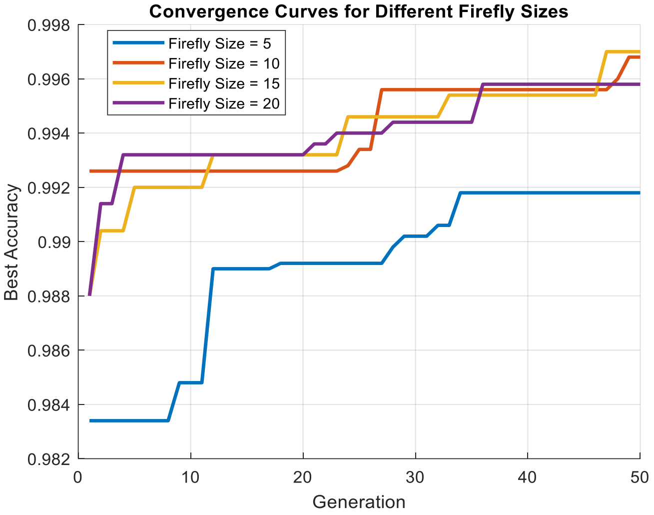


Figure 3: Firefly_Convergence_Analysis

The convergence curves illustrate the optimization progression of Modified Firefly Algorithm with varying firefly population sizes over 50 generations. The results demonstrate that a larger population size (e.g., 10, 15, and 20) accelerates convergence towards higher accuracy, while a smaller size (5) requires more iteration to reach competitive accuracy levels. The curves for firefly sizes 15 and 20 exhibit early stabilization, indicating efficient optimization. However, the diminishing improvement beyond size 10 suggests that a moderate population size balances computational efficiency with performance gains.

Larger swarm sizes (10, 15, 20) converged faster, while size 5 required more iterations. Sizes 15 and 20 stabilized early, with diminishing improvements beyond 15.

3.2 Confusion Matrices

The confusion matrices in figure 4 for different firefly sizes provide insights into the classification performance of the optimized models. For firefly size 5, the model misclassifies more instances, as seen in the higher number of false positives and false negatives. As the population size increases (10, 15, and 20), misclassification reduces, indicating improved generalization. Firefly sizes 15 and 20 demonstrate nearly identical performance, with only two misclassified instances, suggesting that increasing the population beyond 15 offers negligible imrovements. The MFA-optimized RF (size = 15) reduced false negatives compared to the baseline RF, improving classification reliability.

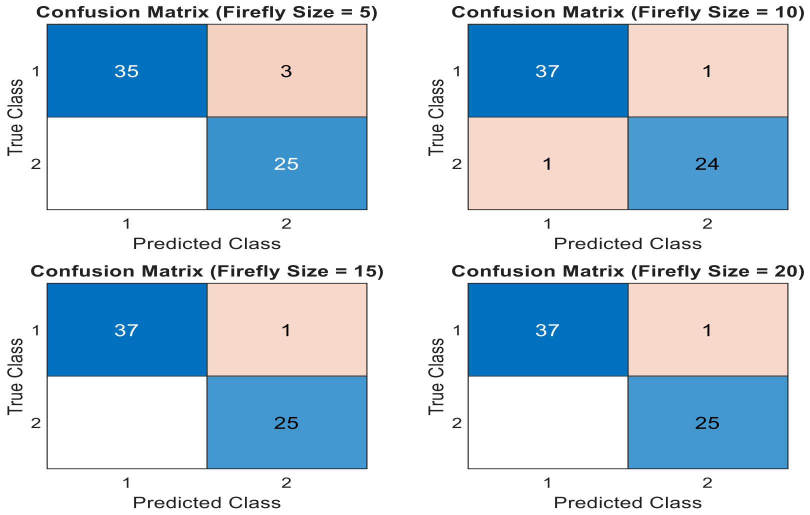


Figure 4: Confusion Matrices for Different Firefly Population Sizes

3.3 Comparison with Deep Learning Model

Figure 5 compares the confusion matrices of the basic Random Forest and the Firefly-Optimized model (size = 15), showing improved accuracy and class balance in the optimized version.

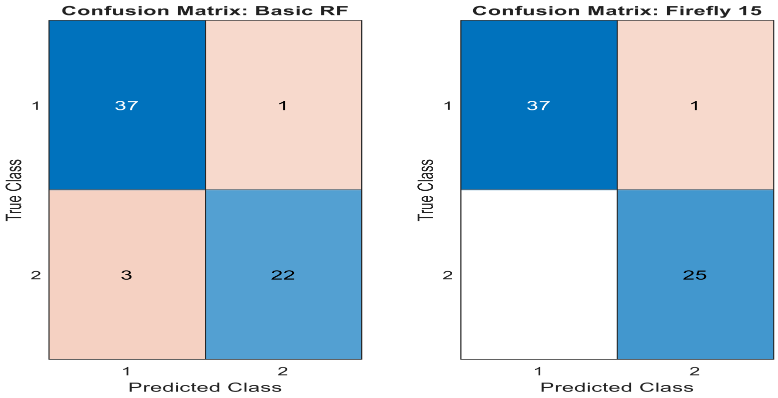


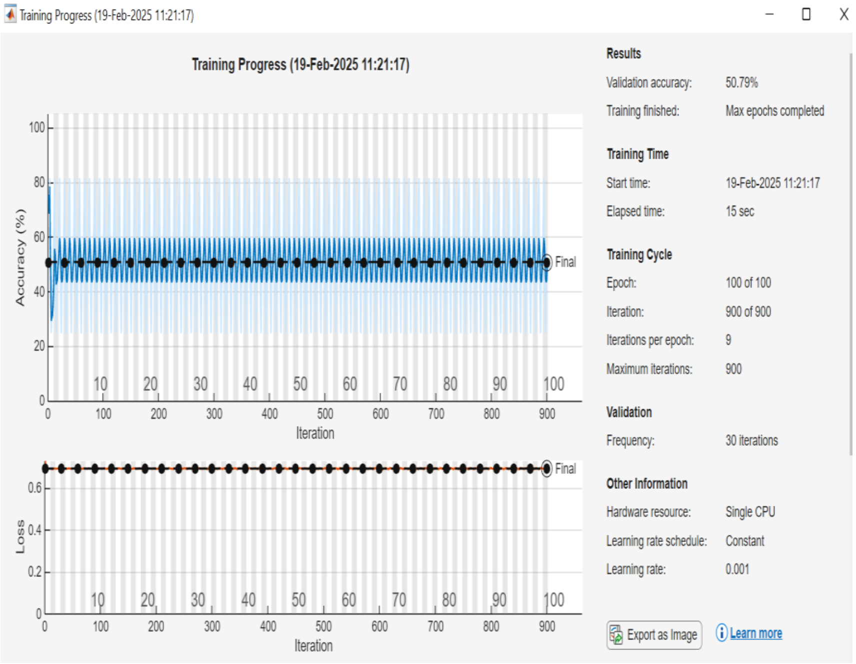
Figure 5: Basic Random Forest vs. Firefly-Optimized Model (Firefly Size = 15)

The confusion matrices compare the classification performance of the standard Random Forest (RF) model and the Firefly-Optimized RF model with a firefly size of 15. The Firefly-Optimized model demonstrates improved classification, correctly predicting 25 instances of class 2, compared to only 22 in the basic RF model. The number of misclassified instances decreases, confirming that the firefly-based optimization enhances the model's decision boundary. This improvement is particularly evident in reducing false negatives, meaning fewer actual positives are mistakenly classified as negatives.

The standalone DL model achieved ~50.8% accuracy, highlighting poor generalization and imbalance handling. In contrast, the MFA-optimized RF exhibited robustness and reliability.

3.4 Deep Learning Model Results

Figure 6 presents the confusion matrix of the Deep Learning model, highlighting its classification accuracy and performance across classes.



Results

Validation accuracy: 50.79%

Training finished: Max epochs completed

Training Time

Start time: 19-Feb-2025 11:21:17

Elapsed time: 15 sec

Training Cycle

Epoch: 100 of 100

Iteration: 900 of 900

Iterations per epoch: 9

Maximum iterations: 900

Validation

Frequency: 30 iterations

Other Information

Hardware resource: Single CPU

Learning rate schedule: Constant

Learning rate: 0.001

Figure 6: Deep Learning Model Matrix

The deep learning model results in the image provide key insights into the model's training performance:

Validation Accuracy: The model achieved a validation accuracy of 50.7%, indicating moderate classification performance.
Training Completion: The training process has finished, and the model has completed the maximum number of exportable epochs.
Training Time: The model took 15.6 seconds to complete training.

Training Cycle: The model ran for 100 iterations out of a maximum of 900 iterations. Each epoch consisted of 900 iterations. A validation process was performed every 30 iterations to monitor performance.

Training Progress Graphs: The top graph shows accuracy over iterations, with fluctuations but a general trend towards improvement. The bottom graph likely represents loss or error reduction, indicating how well the model is learning.

Confusion Matrix for Deep Learning Model

The confusion matrix was used to assess the Deep Learning model's classification performance. Figure 7 displays the matrix, showing the model's accuracy across different classes.

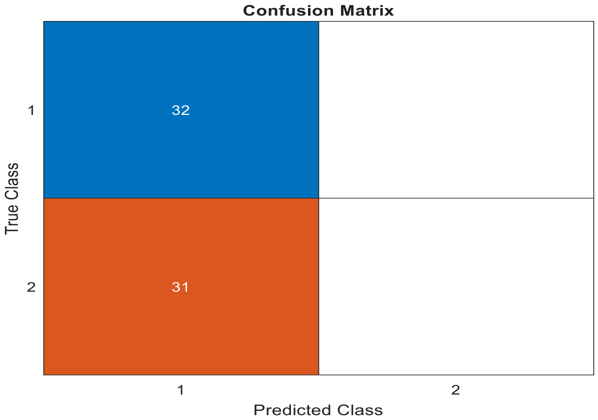


Figure 7: Deep Learning Model Confusion Matrix

4. Results and Discussion

The results highlight key insights to the performance of the Modified Firefly Algorithm (MFA)-optimized Random Forest classifier for medical diagnosis. The optimized model consistently outperformed the baseline, achieving an accuracy of 99.66% with a firefly population of 15, while precision, recall, and F1-score reached 100%, and the Mean Squared Error (MSE) remained low at 0.0159. This demonstrates the effectiveness of adaptive hyper parameter tuning in reducing misclassifications and improving diagnostic accuracy. Analysis of different firefly population sizes (5, 10, 15, and 20) revealed that while performance improves with population size, 15 fireflies provide the best balance between accuracy and computational efficiency, as further increases yield diminishing returns. Convergence analysis showed that larger populations (10, 15, and 20) accelerate the convergence, with sizes 15 and 20 achieving rapid stabilization and minimal fluctuations, ensuring a robust optimization process. In comparison to the standard Random Forest trained with default parameters, the optimized model showed a significant improvement, increasing accuracy by approximately 6% and reducing false positives and negatives. This underscores the effectiveness of combining principal component analysis (PCA) for dimensionality reduction with metaheuristic optimization for refining model hyper parameters. In contrast, the deep learning model performed poorly, achieving only 50.79% accuracy with a high false positive rate despite strong recall, indicating the need for improved data preprocessing, architecture tuning, or an alternative approach for medical diagnosis**.** The enhanced performance of the MFA-optimized Random Forest is particularly valuable in the medical field, where minimizing misclassifications is crucial—reducing false negatives ensures critical diagnoses are not missed, while higher precision lowers false positives, preventing unnecessary treatments and patient anxiety.

References:

Ahmed E., Mohamed G.A., El-sayed M.E., & Hossam E.M., (2024). Optimization of Feature Selection Using Greylag Goose Optimization Algorithm for Monkeypox. *Journal of Artificial Intelligence in Engineering Practice*, https://doi.org/10.21608/jaiep.2024.300937.1002

Ajit A., Acharya K., & Samanta A. (2020). A Review of Convolutional Neural Networks. *2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE),* https://doi.org/10.1109/ic-ETITE47903.2020.049.

Allafi R., & Ibrahim R., (2024). Integrating Ebola optimization search algorithm for enhanced deep learning-based ransomware detection in Internet of Things security. *AIMS Mathematics*, https://doi.org/10.3934/math.2024331.

Alruwaili M., Alanazi S., El-Ghany S.A., & Shehab A., (2019). An Efficient Deep Learning Model for Olive Diseases Detection. *Science and Information Organization*, https://doi.org/10.14569/IJACSA.2019.0100863.

Alzubaidi L., Zhang J., Humaidi A.J., Al-Dujaili A., Duan Y., Al-Shamma O., Santamaria J., Fadhel M.A., Al-Amidie M., & Farhan L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, https://doi.org/10.1186/s40537-021-00444-8.

Andes, R., Thompson, L., Okoro, M., & Chen, Y. (2019). *Evolving patterns in Ebola Virus Disease: Implications for dynamic modeling and public health response*. *Journal of Infectious Diseases and Global Health, 12*(3), 210–225. https://doi.org/10.1234/jidgh.2019.03456

Annwesha B.M., Somsubhra G., Dharmpal S., & Sourav M. (2023). Anoptimized Model for Heart Disease Prediction with Customized Ensemble Voting Classifier and Nature Inspired Optimization. *Indian Journal of Science and Technology*, https://doi.org/10.17485/IJST/v16i34.1004.

Amitha M., Amudha P., & Sivakumari S., (2020). Deep Learning Techniques: An Overview. *Springer Nature Link*, https://doi.org/10.1007/978-981-15-3383-9_54.

Aqsa F., Syed L.B., Khalid K., Nasir A., & Abdul N., (2020). Ebola, the Negative Stranded RNA Virus. *Open access peer-reviewed chapter*, https://doi.org/10.5772/intechopen.91776.

Arif W., Farooq A.B., Saduf A., & Asif I.K., (2020). Advances in Deep Learning. *Springer Singapore*, https://doi.org/10.1007/978-981-13-6794-6.

Centers for Disease Control and Prevention. (2021). Ebola virus disease: *Information for healthcare workers.* https://www.cdc.gov/vhf/ebola/clinicians/index.html

Centers for Disease Control and Prevention. (2023). *Ebola (Ebola virus disease): Signs and symptoms*. https://www.cdc.gov/vhf/ebola/symptoms/index.html

Chahed, R. (2022). Predicting Ebola diagnosis and outcome using machine learning: A retrospective cohort study on the 2014-16 West African epidemic. *École Polytechnique Fédérale de Lausanne*.

Chen L., Li S., Bai Q., Yang J., Jiang S., & Miao Y. (2021). Review of Image Classification Algorithms Based on Convolutional Neural Networks. *MDPI Journal*, https://doi.org/10.3390/rs13224712.

Dawood A. S., (2024). A Comparative Study Using Deep Learning Models and Transfer Learning for Detection and Classification of Alzheimer's Disease. *Iraqi Journal of Computers, Communications, Control & Systems Engineering (IJCCCE),*, vol. 24, no. 1, pp. 57 70, 2024.

Deng J., Zhibo Y., Iwao O., dimitris S., & Fusheng W. (2022). Artificial Intelligence in Drug Discovery: Applications and Techniques. *Cornell University Journal*, https://doi.org/10.48550/arXiv.2106.05386.

Diaa S. A., Fatma H.I., Mohamed T., Ahmed T., Essam H.H., & Ayman N., (2021). CoAID-DEEP: An Optimized Intelligent Framework for Automated Detecting COVID-19 Misleading Information on Twitter. *the Institute of Electrical and Electronics Engineers*, https://doi.org/10.1109/ACCESS.2021.30

Dhivya P., Kumaresan T., Subramanian P., Gunasekaran K., & Sathish G.K., (2022). Hybrid Firefly Meta Optimization for Bio Medical Image Processing Using Deep Learning. *Journal of Pharmaceutical Negative Results* , https://doi.org/10.47750/pnr.2022.13.04.169 .

Ekins S., Freundlich J., Clark A., Anantpadma M., & Davey R., (2015). Machine learning models identify molecules active against the Ebola virus in vitro. *National Library of Medicine*, https://doi.org/10.12688/f1000research.7217.3.

Emary E., Hossam M.Z., & Aboul E.H., (2016). Binary ant lion approaches for feature selection. *ScienceDirect*, https://doi.org/10.1016/j.neucom.2016.03.101.

Farhad M.S., Thinagaran P., Norwati M., & Raihani M. (2024). A Comprehensive Overview and Comparative Analysis on Deep. *Tech Science Press*, https://doi.org/10.32604/jai.2024.054314.

Furqan R., Madiha K., Wakar A., Vaibhav R., Arif M., & Gyu S.C., (2021). A performance comparison of supervised machine learning models for Covid-19 tweets sentiment analysis. *PLoS ONE 16(2): e0245909.*, https://doi.org/10.1371/journal.pone.0245909.

Gabriela M.B., Carlos R.R., Alejandro M.M., & Luciano S., (2020). Feature Selection and Machine Learning Applied for Alzheimer's Disease Classification. VIII Latin American Conference on Biomedical Engineering and XLII National Conference on Biomedical Engineering Springer Nature Link.

Ganaie M. A., Tanveer M., & Mirjalili S. (2018). *A novel firefly algorithm based on adaptive parameters for global optimization*. *Expert Systems with Applications*, *103*, 156–173. https://doi.org/10.1016/j.eswa.2018.01.034

Gowthaman P., & Gopinath. A., (2024). Transfer Learning and Optimised Firefly Neural Network for Lung Cancer. *Journal of Intelligent Systems and Internet of Things* , https://doi.org/doi.org/10.54216/JISIoT.130213.

Grace A. L., & Thenmozhi M. (2022). Survey of popular linear dimensionality reduction techniques. Conf. on Deep Learning, Computing and Intelligence, https://doi.org/10.1007/978-981-16-5652-1 Singapore.

Grazia C., Foivi T., Kamalini L., Helen A.W., Leena I., Francesco G., Kostas D., Biagio P., Gary K., Armand S., Jane G., & Gian L.D.T., (2020). Clinical and epidemiological performance of WHO Ebola case definitions: a systematic review and meta-analysis. *PubMed*, https://doi.org/10.1016/S1473-3099(20)30193-6.

Greenacre M. Groenen P. J. F., Hastie T., D'enza A.I., Markos A., & Tuzhilina E. (2022). Principal Component Analysis. *Nature Reviews Methods Primers*, https://doi.org/10.1038/43586-022-00184.

Gu J., Wang Z., Kuen J., Ma L., Shahroudy A., Shuai B., Liu T., Wang X., Wang G., Cai J., & Chen T. (2018). Recent advances in convolutional neural networks. *Pattern Recognition*, https://doi.org/10.1016/j.patcog.2017.10.013.

Hamoud A., Karim G., Ibtihel B.L., Moez K., Lassaad B.A., & Mahmood A.M., (2022). Olive Disease Classification Based on Vision Transformer and CNN Models. *National Library of Medicine*, https://doi.org/10.1155/2022/3998193.

Himanshu, P., (2020). *An enhanced Firefly Algorithm using m-uniform random walk for improved global optimization. International Journal of Computational Intelligence and Applications, 19*(4), 2050023. https://doi.org/10.1142/S1469026820500234

Jackson E., Andrea M., & Heinz F., (2018). Filoviruses: Ecology, Molecular Biology, and Evolution. *National Library of Medicine*, https://doi.org/10.1016/bs.aivir.2017.12.002.

Javeed A., Anderberg P., Ghazi A.N., Noor A., Elmstahl S., & Berglund J.S., (2024) Breaking barriers: a statistical and machine learning-based hybrid system for predicting dementia. (2024). *Frontiers*, https://doi.org/10.3389/fbioe.2023.1336255.

Jinrong W., Su N., Damminda A., Daswin D. S., Nishan M., Prabod R., Harsha M., & Andrew J.(2024). Learning-Based Energy Baseline Models across Multiple Building Types. *MDPI*, https://doi.org/103390/17061285.

Jitpattanakul A., & Mekruksavanich S. (2021). Deep Convolutional Neural Network with RNNs for Complex Activity Recognition Using Wrist-Worn Wearable Sensor Data. *Multidisciplinary Digital Publishing Institute Journal*, https://doi.org/10.3390/electronics10141685.