Journal of Artificial Intelligence and Metaheuristic

ISSN: 2833-5597 Vol: 11, No. 01, 2026 PP. 32–36

Article DOI: https://doi.org/10.54216/JAIM.110103

FULL LENGTH ARTICLE

A Metaheuristic-Optimized Deep Learning Framework for Accurate Classification of Obsessive–Compulsive Disorder Using Clinical Data Based on the Ninja Optimization Algorithm

Safaa Zaman[1,*] **El-Sayed M. El-Kenawy**[2,3]

[1] Information Sciences Department, College of Life Sciences, Kuwait University, Kuwait

[2] Department of Communications and Electronics, Delta Higher Institute of Engineering and Technology, Mansoura 35111, Egypt

[3] Jadara Research Center, Jadara University, Irbid 21110, Jordan

Emails: S.3zaman@ku.edu.kw · skenawy@ieee.org

Received: August 25, 2025 Revised: October 30, 2025 Accepted: December 24, 2025 ⋆ Corresponding author

ABSTRACT

The growing prevalence and clinical complexity of Obsessive–Compulsive Disorder (OCD) motivate the need for reliable, data-driven decision-support systems capable of improving diagnostic accuracy and robustness beyond traditional assessment methods. This study proposes an optimized deep learning framework that integrates a Deep Learning framework distilled by Gradient Boosting Decision Trees (DeepGBM) with the Ninja Optimization Algorithm (NiOA) to enhance OCD-related classification using structured demographic and clinical data. NiOA is employed for automated hyperparameter tuning of DeepGBM, and the proposed framework is compared against baseline deep learning models and alternative metaheuristic optimizers, including Multiverse Optimization (MVO), Bat Algorithm (BA), and Particle Swarm Optimization (PSO). Experimental evaluation shows that baseline DeepGBM outperforms ANN, CNN, and BiLSTM, achieving an accuracy of 0.8970 and an F-score of 0.8935. Following optimization, the proposed NiOA+DeepGBM framework reaches an accuracy of 0.9779, sensitivity of 0.9763, specificity of 0.9793, and an F-score of 0.9770, consistently surpassing MVO+DeepGBM, BA+DeepGBM, and PSO+DeepGBM across all evaluation metrics. These results confirm the ability of NiOA to navigate complex hyperparameter spaces and improve predictive accuracy and generalization. The proposed model offers a robust clinical decision-support tool for scalable AI-driven mental healthcare systems.

Keywords: Obsessive–Compulsive Disorder ▪ DeepGBM ▪ Ninja Optimization Algorithm ▪ Metaheuristic Optimization ▪ Clinical Decision Support Systems

1. INTRODUCTION

Mental health disorders constitute a major global health concern, exerting profound and long-lasting effects on individuals, healthcare systems, and societies. Within this landscape, Obsessive–Compulsive Disorder (OCD) is particularly challenging because of its chronic nature, symptom variability, and frequent comorbidity with other mental health disorders. The integration of data-driven methodologies, especially machine learning and deep learning, provides promising opportunities for improving clinical decision-making in OCD assessment.

OCD is characterized by persistent intrusive thoughts (obsessions) and repetitive behaviours or mental acts (compulsions) performed to reduce distress. The disorder often follows a chronic course, with symptoms fluctuating in severity over time, and is frequently associated with functional impairment, reduced quality of life, and elevated risk of depression and anxiety. From a societal perspective, OCD imposes a substantial economic burden through healthcare use, productivity loss, and long-term disability.

Accurate and timely diagnosis remains difficult in clinical practice. Traditional approaches rely heavily on clinical interviews and standardized rating scales, which are valuable but subjective and dependent on clinician expertise. Predictive modelling can augment conventional assessment by systematically analysing patient data and supporting evidence-based decision-making.

This paper develops a unified framework in which DeepGBM is optimized using NiOA. The key contributions are: (i) a clinical-data-driven OCD classification pipeline; (ii) automated hyperparameter tuning using NiOA; (iii) comparison with ANN, CNN, BiLSTM, and DeepGBM baselines; and (iv) comparison with MVO, BA, and PSO optimized variants under the same evaluation protocol.

2. LITERATURE REVIEW

Research on OCD increasingly uses machine learning and deep learning to address limitations of conventional clinical assessment. Neuroimaging-based models have demonstrated the potential of structural MRI, functional MRI, resting-state connectivity, and EEG-derived biomarkers for identifying OCD-related patterns. Other studies have used demographic, behavioural, and symptom-level data to predict severity, remission, comorbid depression, and long-term clinical course.

Deep learning architectures are attractive because they can model nonlinear interactions among heterogeneous predictors. ANN, CNN, recurrent models, and hybrid architectures have been applied to psychiatric prediction tasks; however, these models are often sensitive to hyperparameter settings and may suffer from overfitting when data are limited. Ensemble and gradient-boosting approaches provide strong tabular-data performance, while model distillation strategies can combine representation learning with decision-tree interpretability and robustness.

Metaheuristic optimization algorithms are increasingly used for feature selection and hyperparameter tuning. Population-based algorithms such as PSO, MVO, BA, and NiOA are useful when objective functions are nonlinear, multimodal, and not analytically differentiable. Their role is particularly important in clinical decision-support systems, where robust generalization is more important than merely minimizing training error.

3. METHODOLOGY

The proposed methodology combines structured clinical data preprocessing, baseline deep learning classification, DeepGBM modelling, and metaheuristic optimization. The framework is designed to classify OCD-related clinical records while improving generalization through automated hyperparameter search.

3.1 Problem Formulation

Let $\mathbf{X} \in \mathbb{R}^{m\times n}$ denote a clinical dataset containing m patient records and n demographic or clinical attributes, and let $\mathbf{y} \in \{0,1\}^m$ denote the target class labels. The task is to learn a classifier

$$f : \mathbf{X} \to \mathbf{y} \tag{1}$$

that maximizes predictive performance on unseen patient records while maintaining robustness across sensitivity, specificity, and predictive-value metrics.

3.2 DeepGBM Classification Model

DeepGBM combines deep representation learning with Gradient Boosting Decision Trees. The deep component maps each input vector $\mathbf{x}_i$ into a latent representation:

$$\mathbf{h}_i = \phi(\mathbf{W}\mathbf{x}_i + \mathbf{b}), \tag{2}$$

where $\phi(\cdot)$ is a nonlinear activation function, $\mathbf{W}$ is a weight matrix, and $\mathbf{b}$ is a bias vector. The gradient boosting component then learns an additive model:

$$F_M(\mathbf{x}) = \sum_{m=1}^{M} \eta f_m(\mathbf{x}), \tag{3}$$

where M is the number of trees, f_m is the mth weak learner, and η is the learning rate.

3.3 Ninja Optimization Algorithm

NiOA is used as the main hyperparameter optimization strategy. Each candidate solution represents a set of DeepGBM hyperparameters, including learning rate, number of estimators, tree depth, regularization terms, and neural representation parameters. The optimization objective is formulated as

$$\mathbf{z}^* = \arg\max_{\mathbf{z}\in\Omega} \mathscr{F}\big(\text{DeepGBM}(\mathbf{z})\big), \tag{4}$$

where Ω is the search space and $\mathscr{F}$ is the validation performance criterion.

3.4 NiOA Pseudocode

The NiOA procedure used in the study can be summarized as follows.

1. Initialize the population size N, maximum iterations T, candidate positions, control parameters, mutation parameter, velocity factor, and best solution B_s.
2. Set the iteration counter $t = 0$.
3. Evaluate the fitness of all candidate hyperparameter vectors using validation performance.
4. Update candidate positions through exploration and exploitation operators.
5. Apply mutation and stagnation-control mechanisms to avoid premature convergence.
6. Update the best solution whenever a candidate improves the objective value.
7. Repeat until $t = T$, then return the best DeepGBM hyperparameter vector.

4. EXPERIMENTAL SETUP

The experimental design evaluates baseline deep learning models and optimized DeepGBM variants on OCD-related structured clinical data. The compared baseline models are ANN, CNN, BiLSTM, and DeepGBM. The optimized models combine DeepGBM with NiOA, MVO, BA, and PSO. All models are evaluated using accuracy, sensitivity, specificity, positive predictive value, negative predictive value, and F-score. Figures 1–3 present the principal exploratory data visualizations reported before model evaluation.

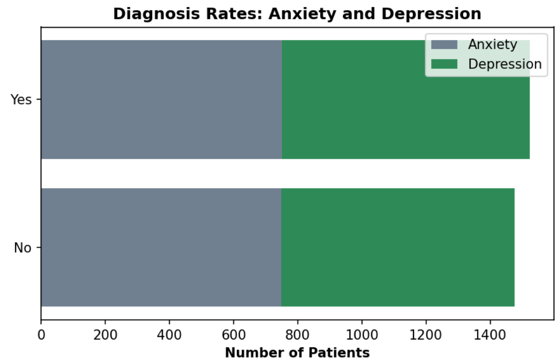

Figure 1. Distribution of patients with and without clinical diagnoses of anxiety and depression.

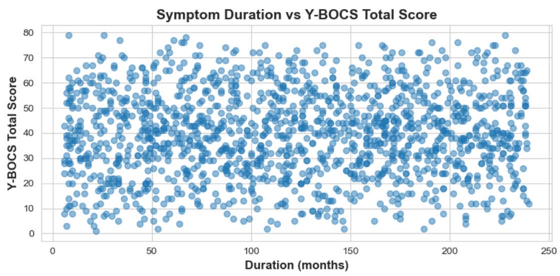

Figure 2. Scatter plot illustrating the relationship between symptom duration and Y-BOCS total score.

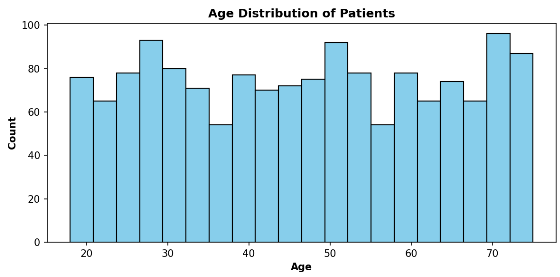

Figure 3. Histogram illustrating the age distribution of patients included in the study.

4.1 Evaluation Metrics

Table 1 gives the formal metric definitions used throughout the experimental analysis.

Table 1. Classification evaluation metrics and mathematical definitions.

Metric	Mathematical equation
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
Sensitivity (TPR)	$\frac{TP}{TP+FN}$
Specificity (TNR)	$\frac{TN}{TN+FP}$
PPV	$\frac{TP}{TP+FP}$
NPV	$\frac{TN}{TN+FN}$
F-Score	$\frac{2TP}{2TP+FP+FN}$

5. RESULTS AND DISCUSSION

This section presents the baseline and optimized classification results. The baseline stage establishes the relative strength of DeepGBM, while the optimization stage evaluates the impact of metaheuristic tuning.

5.1 Baseline Deep Learning Results

DeepGBM achieves the strongest baseline results, with accuracy of 0.8970 and F-score of 0.8935. BiLSTM ranks second, followed by CNN and ANN. Table 2 summarizes the reported baseline results, while Figures 4–6 visualize the same baseline comparisons.

Table 2. Baseline performance of deep learning models before optimization.

Model	Accuracy	Sensitivity	Specificity	PPV	NPV	F-Score
DeepGBM	0.897009967	0.890410959	0.903225806	0.896551724	0.897435897	0.893470790
BiLSTM	0.885135135	0.875000000	0.894736842	0.887323944	0.883116883	0.881118881
CNN	0.872727273	0.861111111	0.883720930	0.873239437	0.872093023	0.867132867
ANN	0.859060403	0.849315068	0.868279570	0.861111111	0.857142857	0.855172414

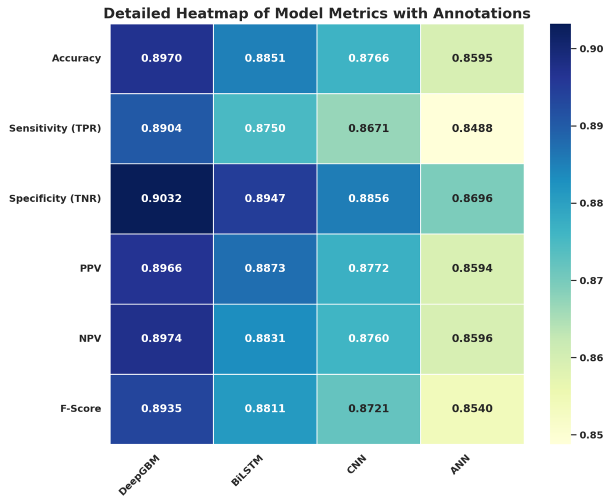

Figure 4. Heatmap visualization of performance metrics for four classification models: DeepGBM, BiLSTM, CNN, and ANN.

5.2 Optimized DeepGBM Results

After hyperparameter optimization, NiOA+DeepGBM obtains the highest performance across all metrics. It achieves an accuracy of 0.977851463, sensitivity of 0.976301736, specificity of 0.979294908, PPV of 0.977737665, NPV of 0.977957156, and F-score of 0.977019173. These values demonstrate that NiOA improves both positive-class detection and negative-class discrimination. Table 3 reports the optimized scores.

Table 3. Performance comparison of optimized DeepGBM models using different metaheuristic algorithms.

Model	Accuracy	Sensitivity	Specificity	PPV	NPV	F-Score
NiOA + DeepGBM	0.977851463	0.976301736	0.979294908	0.977737665	0.977957156	0.977019173
MVO + DeepGBM	0.967023521	0.962031811	0.971223021	0.968859649	0.965812000	0.965433403
BA + DeepGBM	0.953961973	0.949275362	0.958762887	0.955056180	0.953846154	0.952156057
PSO + DeepGBM	0.944259568	0.938405797	0.950354610	0.944525547	0.944444444	0.941455697

After hyperparameter optimization, Figures 7–9 provide the corresponding visual comparison for the DeepGBM variants optimized by NiOA, MVO, BA, and PSO.

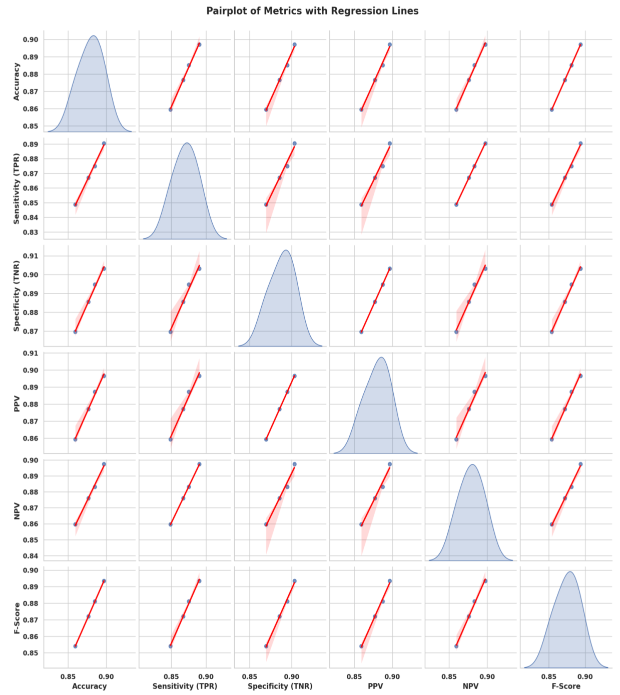

Figure 5. Pairplot visualization of baseline classification performance metrics, including accuracy, sensitivity, specificity, PPV, NPV, and F-score.

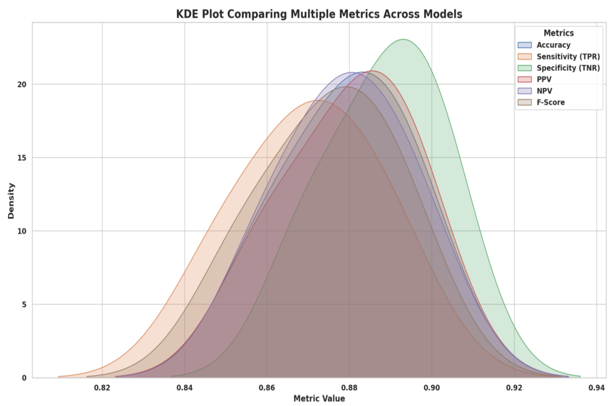

Figure 6. Kernel density estimation plots comparing the distributions of baseline classification performance metrics.

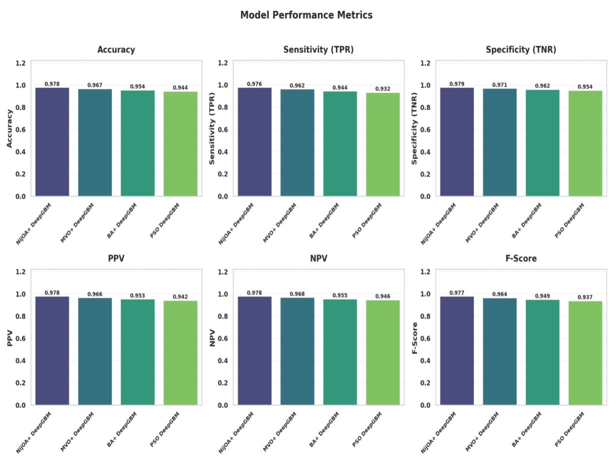

Figure 7. Bar chart comparison of classification performance metrics for four hybrid DeepGBM models.

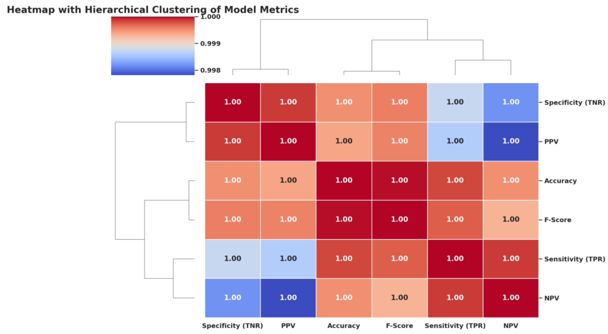

Figure 8. Correlation heatmap of optimized classification performance metrics with hierarchical clustering.

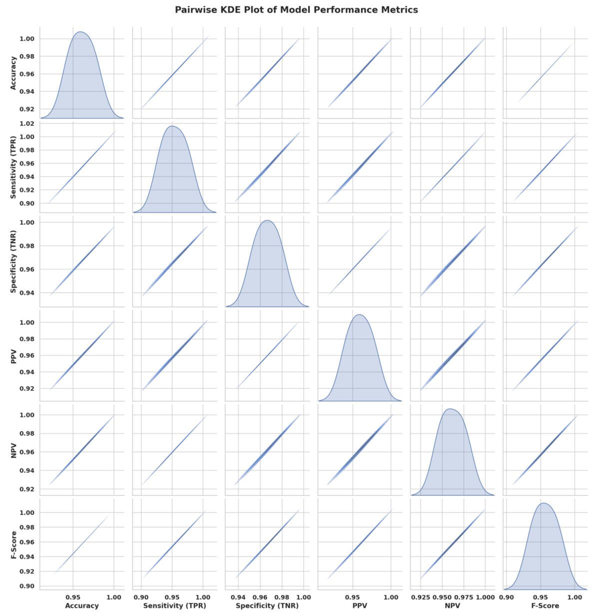

Figure 9. Pairwise kernel density estimation plot of optimized model performance metrics.

The improvement produced by NiOA is clinically meaningful. In psychiatric applications, false negatives may delay treatment and false positives may lead to unnecessary interventions. The simultaneous improvement in sensitivity and specificity therefore indicates that the optimized model is better balanced than baseline and alternative optimized variants. The results also suggest that NiOA provides a stronger exploration–exploitation balance in the complex hyperparameter landscape of DeepGBM.

6. CONCLUSION

This paper presents a metaheuristic-optimized deep learning framework for OCD classification using structured clinical data. The proposed NiOA+DeepGBM model substantially improves over baseline deep learning models and alternative optimized variants. Baseline DeepGBM achieves an accuracy of 0.8970 and F-score of 0.8935, while NiOA-optimized DeepGBM reaches 0.9779 accuracy and 0.9770 F-score. Sensitivity and specificity also improve to 0.9763 and 0.9793, respectively.

The findings indicate that integrating NiOA with DeepGBM provides an effective, robust, and clinically relevant decision-support model. Future work should evaluate the framework on larger multi-centre datasets, include multimodal sources such as neuroimaging and electronic health records, and investigate deployment constraints for real-time mental healthcare applications.

REFERENCES

[1] World Health Organization, *World Mental Health Report: Transforming Mental Health for All*. Geneva, Switzerland: WHO, 2022.

[2] B. Stahnke, “A systematic review of misdiagnosis in those with obsessive-compulsive disorder,” *Journal of Affective Disorders Reports*, vol. 6, p. 100231, 2021.

[3] F.-F. Huang et al., “Functional and structural MRI based obsessive-compulsive disorder diagnosis using machine learning methods,” *BMC Psychiatry*, vol. 23, no. 1, p. 792, 2023.

[4] B. A. Zaboski and L. Bednarek, “Precision psychiatry for obsessive-compulsive disorder: Clinical applications of deep learning architectures,” *Journal of Clinical Medicine*, vol. 14, no. 7, p. 2442, 2025.

[5] B. A. Zaboski, A. Wilens, J. P. McNamara, and G. N. Muller, “Predicting OCD severity from religiosity and personality: A machine learning and neural network approach,” *Journal of Mood & Anxiety Disorders*, vol. 8, p. 100089, 2024.

[6] M. Grassi et al., “Prediction of illness remission in patients with obsessive-compulsive disorder with supervised machine learning,” *Journal of Affective Disorders*, vol. 296, pp. 117–125, 2022.

[7] C. Segalàs et al., “Cognitive and clinical predictors of a long-term course in obsessive compulsive disorder: A machine learning approach in a prospective cohort study,” *Journal of Affective Disorders*, vol. 350, pp. 648–655, 2024.

[8] W. B. Bruin et al., “The functional connectome in obsessive-compulsive disorder: Resting-state mega-analysis and machine learning classification for the ENIGMA-OCD consortium,” *Molecular Psychiatry*, vol. 28, no. 10, pp. 4307–4319, 2023.

[9] S. Hinduja et al., “Multimodal prediction of obsessive-compulsive disorder and comorbid depression severity and energy delivered by deep brain electrodes,” *IEEE Transactions on Affective Computing*, vol. 15, no. 4, pp. 2025–2041, 2024.

[10] Z. Zhao et al., “Exploring potential resting-state EEG biomarkers of obsessive-compulsive disorder based on explainable machine learning analysis,” *BMC Psychiatry*, vol. 25, no. 1, p. 1164, 2025.