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

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. In this study, we propose an optimized deep learning framework that integrates a Deep Learning framework distilled by Gradient Boosting Decision Trees (DeepGBM) with a novel metaheuristic optimizer, the Ninja Optimization Algorithm (NiOA), to enhance OCD-related classification using structured demographic and clinical data. The main contribution of this work lies in the design of a unified optimization pipeline in which NiOA is employed for automated hyperparameter tuning of DeepGBM, and in the comprehensive comparison of this approach against baseline deep learning models and alternative metaheuristic optimizers, including Multiverse Optimization (MVO), Bat Algorithm (BA), and Particle Swarm Optimization (PSO). Experimental evaluation demonstrates that, at the baseline stage, DeepGBM outperforms Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), and Bidirectional Long Short-Term Memory networks (BiLSTM), achieving an accuracy of 0.8970 and an F-score of 0.8935. Following optimization, the proposed NiOA+DeepGBM framework achieves substantial performance gains, reaching 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 superior capability of NiOA in navigating complex hyperparameter spaces and enhancing both predictive accuracy and generalization. The implications of this work are significant for intelligent mental health assessment, as the proposed NiOA-optimized DeepGBM model offers a robust, clinically relevant decision-support tool that can assist clinicians in improving diagnostic reliability, reducing uncertainty, and supporting the development of 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 at large [1], [2]. The increasing prevalence of psychiatric conditions, coupled with their complex etiologies and heterogeneous clinical manifestations, has intensified the demand for advanced analytical tools capable of supporting accurate diagnosis, prognosis, and treatment

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

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

planning. Within this broad landscape, Obsessive-Compulsive Disorder (OCD) stands out as a particularly challenging condition due to its chronic nature, symptom variability, and frequent comorbidity with other mental health disorders [3], [4], [5]. The integration of data-driven methodologies, especially machine learning and deep learning techniques, offers promising avenues for addressing these challenges and enhancing clinical decision-making in OCD assessment [6], [7], [8], [9].

Obsessive-Compulsive Disorder is a debilitating psychiatric condition characterized by persistent intrusive thoughts (obsessions) and repetitive behaviors or mental acts (compulsions) performed to alleviate distress. Epidemiological evidence indicates that OCD affects millions of individuals worldwide, cutting across age groups, genders, and cultural backgrounds. The disorder often follows a chronic course, with symptoms fluctuating in severity over time, and it is frequently associated with significant functional impairment, reduced quality of life, and elevated risk of comorbid conditions such as depression and anxiety. From a societal perspective, OCD imposes a substantial economic burden due to increased healthcare utilization, loss of productivity, and long-term disability.

Accurate and timely diagnosis of OCD remains a critical challenge in clinical practice [10], [11], [12]. Delayed identification or misclassification can result in prolonged suffering, inappropriate treatment strategies, and diminished therapeutic outcomes. Traditional diagnostic approaches rely heavily on clinical interviews and standardized rating scales, which, while valuable, are inherently subjective and dependent on clinician expertise. As a result, there is growing interest in augmenting conventional assessment methods with predictive modeling techniques that can systematically analyze large volumes of patient data and support evidence-based decision-making.

Predictive modeling plays an increasingly important role in modern mental healthcare by enabling early detection of psychiatric conditions, stratification of symptom severity, and optimization of treatment pathways. In the context of OCD, data-driven models can assist in identifying complex relationships among demographic characteristics, clinical history, and diagnostic indicators, thereby facilitating more personalized and proactive care. Accurate classification models have the potential to enhance clinical efficiency, reduce diagnostic uncertainty, and contribute to the development of individualized treatment plans tailored to patient-specific profiles [13], [14], [15].

Machine learning and deep learning techniques have emerged as powerful tools for predictive analytics in psychiatry due to their ability to handle high-dimensional and heterogeneous datasets. Unlike traditional statistical methods, these approaches can model non-linear interactions and uncover latent patterns within structured clinical data. Deep learning architectures, in particular, offer advanced representation learning capabilities that are well-suited to complex healthcare datasets. Recent advances in ensemble deep learning frameworks, such as DeepGBM, which integrates neural networks with gradient boosting decision trees, have demonstrated strong potential for structured data analysis by combining predictive accuracy with improved robustness and interpretability. These characteristics make such models especially relevant for OCD-related classification tasks.

1.1 Challenges in OCD Patient Classification

Despite the potential benefits of machine learning-based approaches, OCD patient classification presents several methodological and practical challenges. One of the foremost issues is the high dimensionality and heterogeneity of available clinical data. OCD datasets typically encompass a wide range of demographic variables, diagnostic timelines, symptom duration measures, and psychiatric history indicators. The integration of these diverse features into a unified modeling framework increases computational complexity and may hinder effective learning if not carefully managed.

Feature redundancy and irrelevance further complicate the classification process. Many demographic and clinical variables exhibit strong correlations, and the inclusion of redundant features can introduce noise, inflate model complexity, and degrade predictive performance. Without appropriate feature selection mechanisms, models may struggle to identify the most informative attributes, leading to reduced stability and generalization capability. This challenge is particularly pronounced in deep learning models, where large parameter spaces amplify the impact of irrelevant inputs.

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

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

Another critical challenge lies in the sensitivity of machine learning and deep learning models to hyperparameter configurations. Model performance is often highly dependent on the selection of learning rates, network architectures, regularization parameters, and other hyperparameters. Suboptimal tuning can result in underfitting, overfitting, or inefficient convergence, thereby limiting the practical applicability of the model. Consequently, systematic and automated hyperparameter optimization strategies are essential for achieving reliable performance in OCD classification tasks.

Generalization and overfitting remain central concerns in psychiatric data analysis. Models trained on limited or biased datasets may perform well during training but fail to maintain accuracy when applied to unseen patient data. Ensuring robust generalization across diverse OCD populations requires careful model design, effective feature selection, and rigorous optimization procedures that balance model complexity with predictive reliability.

The primary objective of this study is to develop and evaluate advanced machine learning and deep learning models for OCD patient classification using comprehensive demographic and clinical data. The study aims to conduct a systematic comparative analysis of multiple predictive models, including DeepGBM, Bidirectional Long Short-Term Memory networks, Convolutional Neural Networks, and Artificial Neural Networks, in order to assess their suitability for structured psychiatric datasets.

In addition to baseline model evaluation, the study seeks to enhance predictive performance through the application of metaheuristic optimization techniques. Specifically, optimization algorithms such as the Ninja Optimization Algorithm, Multiverse Optimization, Bat Algorithm, and Particle Swarm Optimization are employed to address challenges related to feature selection and hyperparameter tuning. By integrating these optimization strategies, the study aims to reduce model complexity, improve learning efficiency, and strengthen generalization performance.

The overarching goal is to improve key classification characteristics, including accuracy, sensitivity, specificity, and overall predictive balance, while simultaneously minimizing unnecessary computational overhead. Through this approach, the study aspires to contribute a robust and scalable predictive framework that supports data-driven OCD assessment in clinical settings.

This work contributes to the growing body of literature on machine learning applications in mental health by proposing an optimized DeepGBM-based classification framework tailored to OCD patient data. A key contribution lies in the integration of the Ninja Optimization Algorithm with DeepGBM, enabling joint feature selection and hyperparameter optimization within a unified modeling pipeline.

Furthermore, the study offers a comprehensive comparative perspective by examining multiple machine learning models and optimization strategies under a consistent experimental framework. This analysis provides valuable insights into the relative effectiveness of different metaheuristic algorithms for psychiatric classification tasks.

Finally, the proposed hybrid optimization strategy highlights the synergistic role of feature selection and hyperparameter tuning in enhancing predictive performance and computational efficiency. By addressing both aspects simultaneously, the study advances methodological best practices for developing reliable and interpretable machine learning models in mental healthcare applications.

The remainder of this paper is organized as follows. The next section describes the dataset, preprocessing procedures, machine learning models, and metaheuristic optimization techniques employed in the study. This is followed by a detailed presentation of the experimental design and evaluation methodology. Subsequently, the findings are analyzed and discussed in relation to existing literature and clinical relevance. The paper concludes with a summary of key insights and outlines potential directions for future research in optimized machine learning approaches for OCD assessment.

2 Literature Review

Research on obsessive–compulsive disorder (OCD) increasingly leverages machine learning (ML) and deep learning (DL) to address persistent limitations of conventional clinical assessment, especially the subjectivity

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

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

of interview-based diagnosis and the difficulty of forecasting treatment trajectories in a heterogeneous disorder. Across the provided studies, a consistent theme is the attempt to transform diverse, high-dimensional measurements—neuroimaging, EEG, biochemical assays, gene expression, cognitive testing, behavioral signals, and large-scale survey features—into clinically meaningful predictions. At the same time, the literature repeatedly emphasizes that high reported accuracy in constrained datasets does not automatically translate into reliable generalization across sites, populations, or acquisition pipelines, motivating a growing focus on external validation, standardization, and interpretability.

Neuroimaging-based classification studies illustrate both the promise and the current ceiling of MRI biomarkers for individual-level diagnosis. One investigation using fifty OCD patients and fifty healthy controls extracted multiple resting-state fMRI indices (ALFF, fALFF, ReHo, and DC) and complementary sMRI morphometric measures, reduced features with LASSO, and compared several classifiers [16]. The most successful approach combined multiple fMRI indices within an SVM framework, achieving strong internal cross-validation and improved test-set discrimination relative to models using single indices or combined fMRI+sMRI feature pools. Conceptually, this result suggests that different fMRI metrics capture partially non-overlapping aspects of abnormal intrinsic activity and connectivity, so their integration can amplify disease-relevant signal beyond any single representation [16]. Importantly, the discriminative features were not restricted to canonical cortico-striato-thalamo-cortical circuits but extended to additional brain regions, reinforcing the view that OCD-related neural signatures may be distributed rather than localized, and that model design should avoid overly narrow region-of-interest assumptions [16].

However, large-scale connectivity evidence tempers optimism about resting-state functional connectivity as a standalone diagnostic biomarker. A mega-analysis pooling data from 28 ENIGMA-OCD samples (over one thousand OCD patients and a comparable number of healthy controls) found widespread but small-to-moderate group differences characterized predominantly by hypo-connectivity, with comparatively few hyper-connections (notably involving the thalamus) [17]. Critically, despite the unprecedented sample size and whole-brain scope, the machine learning classification performance remained low overall, with only modest improvements in medicated subgroups [17]. Methodologically, this juxtaposition is instructive: statistical detectability at the group level does not guarantee clinically useful separability at the individual level, particularly when effect sizes are small, scanner/site variability is substantial, and the disorder itself is phenotypically diverse [17]. In practical terms, these findings motivate either richer feature representations (e.g., multimodal integration or task-based measures) or clinically informed stratification strategies, rather than assuming that resting-state connectivity alone can yield robust diagnostic tools.

Structural MRI has also been positioned as a substrate for dimensional prediction, shifting the target from categorical diagnosis to symptom severity estimation. A study applying support vector regression to gray matter volumes in untreated adults asked whether morphometric patterns in cortical–subcortical loops could predict clinical severity. This framing aligns with contemporary psychiatric research that treats symptoms as continuous traits rather than discrete states, potentially yielding models that track clinical change or support stratified treatment planning. At the same time, the study explicitly acknowledges small-sample constraints, underscoring a central challenge in neuroimaging ML: models may appear promising in pilot settings yet require replication in independent cohorts to establish stability, transportability, and resistance to overfitting[18].

Within this landscape, DL is often presented as a toolkit for handling the complexity of psychiatric data rather than a guaranteed performance upgrade. A DL-focused review synthesizes how multiple architectures—from feedforward networks to CNNs, RNNs, GANs, and transformers—can be mapped onto OCD research problems, including diagnosis, outcome prediction, and even data generation to address scarcity [19]. The review emphasizes that architecture choice should follow data structure (e.g., CNNs for spatially organized neuroimaging, RNNs/transformers for temporal sequences) and that DL can support precision psychiatry by learning nonlinear interactions that may be missed by traditional models [19]. Yet, the same account highlights barriers to clinical adoption, particularly interpretability and implementation feasibility, implying that DL success in OCD will depend as much on transparent validation and deployment design as on raw predictive accuracy [19].

A complementary review explicitly foregrounds early detection and argues for hybrid ML–XAI pipelines that integrate neuroimaging (sMRI/fMRI), clinical markers, and biochemical features, while using interpretability methods such as SHAP to increase clinical trust [20]. In this perspective, interpretability is not an optional add-on but a requirement in mental health contexts where decisions must be explainable to clinicians

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

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

and patients, and where models trained on heterogeneous datasets may inadvertently encode confounds. The review also stresses dataset variability, limited generalizability, and ethical concerns, implying that performance claims should be evaluated alongside data governance, privacy, and fairness considerations [20]. Together with DL-oriented discussions, this work positions the field as moving from proof-of-concept classification toward clinically credible systems that can justify outputs, tolerate domain shift, and operate within real-world constraints.

EEG-based approaches further illustrate the tension between feasibility and standardization. A systematic review of EEG-ML classification studies (searching major databases up to February 2025) identified only a small number of eligible studies and reported marked heterogeneity in populations, preprocessing, EEG feature engineering, validation practices, and reporting standards [21]. This heterogeneity limits synthesis and makes it difficult to distinguish genuine biomarker signal from pipeline-specific artifacts. Notably, the review highlights that modern interpretability techniques were absent from the surveyed literature, even though such methods could reduce black-box concerns and potentially guide electrode placement for interventions [21]. This diagnosis of the evidence base suggests that, for EEG-ML in OCD, methodological harmonization and transparent reporting may be as important as algorithm selection.

In contrast, an explainable EEG-ML study directly operationalized these priorities by using two independent datasets, extracting multiple EEG feature sets, optimizing models on one dataset, and testing generalizability on an external set [22]. The results emphasized phase-locking value (PLV) functional connectivity as particularly informative across classifiers, with LightGBM achieving strong performance in both development and external testing [22]. Crucially, SHAP-based interpretation highlighted specific band-limited long-range connections as influential predictors, linking model behavior to neurophysiological hypotheses rather than leaving it as an opaque mapping [22]. This study is important not only for its accuracy but also because it demonstrates a workflow that aligns with clinical expectations: external validation, feature-set comparison under a unified framework, and interpretable attribution of predictive drivers [22].

Beyond brain signals, biochemical frameworks aim to embed OCD prediction into smart healthcare infrastructures. The Accu-Help conceptual model proposes automated OCD detection via oxidative stress biomarkers (OSBs), framing the end-to-end pipeline as a distributed data integration and continuous model updating problem involving hospitals and biochemical laboratories [23]. By defining clinically relevant classes (healthy, OCD-affected, genetically affected) and deploying a neural network prediction module, the study situates ML not merely as an analytic method but as a component of a socio-technical system requiring real-time coordination, dataset growth management, and model redeployment [23]. Although accuracy is reported as high, the most distinctive contribution is arguably the articulation of operational requirements for translation—data labeling logistics, integration workflows, and model availability across sites—which are often under-discussed in purely methodological papers [23].

At the molecular level, gene expression studies attempt to bypass symptom-based diagnosis by identifying biological signatures measurable in peripheral blood or brain tissue. One study sourced expression data from the GEO repository and proposed a hybrid feature selection strategy combining statistical and ML techniques to identify down-regulated genes implicated in OCD, reporting strong classification performance for both blood and brain datasets compared with alternative feature selection methods [24]. The clinical relevance of this direction is twofold: first, it suggests potential for objective molecular screening or risk stratification; second, it aligns with precision medicine logic in which expression profiles can help identify receptor-level targets and inform individualized treatment hypotheses [24]. Nevertheless, as with other modalities, the implied requirement is careful validation across cohorts, since transcriptomic signals may be sensitive to tissue source, batch effects, medication status, and comorbidities.

A major clinical objective in OCD is not only diagnosis but also prediction of treatment response and longitudinal outcome. One study used cognitive and clinical variables to predict pharmacological response over 12 weeks, drawing on standardized symptom scales (including Y-BOCS) and executive functioning measures such as working memory and abstract reasoning [25]. The reported high correspondence between predicted and observed outcomes suggests that cognitive profiles can add predictive value beyond baseline symptom severity alone, consistent with the clinical observation that preserved executive functioning may relate to better medication response [25]. From a modeling standpoint, the key implication is that routinely collected cognitive and demographic data can support personalized treatment planning, potentially reducing trial-and-error prescribing by identifying likely responders to a given strategy [25].

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

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

Long-term outcomes pose an even harder problem because they reflect illness chronicity, treatment history, and evolving psychosocial context. A study following outpatients for 5–17 years used baseline demographic and neuropsychological data, as well as early pharmacological response, to predict long-term Y-BOCS severity with ML [26]. The finding that adding early response improved prediction is clinically intuitive: early trajectory can function as a summary statistic capturing latent severity, treatment sensitivity, and adherence dynamics. The reported high proportion of non-responders at long-term follow-up also frames the prediction task as high-stakes, since accurate prognosis could influence escalation decisions, including psychotherapy intensification or more invasive options in severe cases [26]. While limitations such as sample size and severity-biased recruitment are noted, the study illustrates how ML can be used to formalize prognostic reasoning and quantify expected outcomes [26].

A separate multi-center effort aimed to predict remission after two years using only predictors accessible in routine clinical care, employing gradient-boosted decision trees with cross-validated Bayesian hyperparameter optimization [27]. Moderate performance in independent centers, coupled with significant between-center variation, highlights a central translational barrier: even when assessment protocols are harmonized, models can behave differently across clinical settings due to subtle shifts in patient mix, clinician practices, or measurement distributions [27]. The identification of predictors related to severity, chronic course, medication use, and global functioning reinforces that remission is multifactorial and not reducible to a single biomarker axis [27]. This study thus advances the field by explicitly testing portability across sites and by diagnosing instability as a target for future methodological improvement [27].

Psychological and behavioral modeling further broadens the conception of OCD prediction targets beyond diagnosis. A study examining OCD severity in relation to personality traits, religiosity, spirituality, and demographic factors found that item-level features could be more informative than aggregated scale scores [28]. This suggests that conventional psychometric summarization may discard predictive nuance, whereas ML can preserve fine-grained patterns and nonlinear interactions that better reflect heterogeneity. Interestingly, the work also noted that a neural network did not necessarily outperform linear regression in accuracy, yet was valuable for capturing nonlinear relationships and supporting richer interpretive hypotheses about OCD severity drivers [28]. The broader methodological lesson is that model evaluation should consider both predictive performance and the type of structure a model can reveal about complex psychological phenomena [28].

Population-scale screening approaches, particularly for youth mental health, leverage large feature spaces derived from survey data to support early recognition. One study used 1,474 features from a nationally representative Australian child and adolescent mental health survey and identified algorithms that performed strongly for OCD classification, reporting high accuracy and specificity [29]. The development of a Streamlit-based application emphasizes translation into accessible tools for parents/guardians and school officials, though the broader research implication is that symptom-and-context feature breadth can compensate for limitations of smaller clinical datasets, provided that validation and responsible deployment are addressed [29]. Such approaches also raise important methodological considerations about feature redundancy, selection bias, and the interpretability of high-dimensional symptom questionnaires in real-world settings [29].

Finally, multimodal behavioral sensing and neuromodulation-focused modeling demonstrate an advanced frontier of outcome measurement and intervention control. In a clinical trial of deep brain stimulation for refractory OCD, random forest regression models were trained on multimodal audiovisual and speech-derived features to predict OCD severity, comorbid depression severity, and total electrical energy delivered, with Shapley-based feature reduction enhancing efficiency and predictive power. The finding that multimodal measures outperformed single-modality inputs suggests that OCD-related affective and symptomatic expression is better captured through integrated behavioral channels rather than isolated signals. Moreover, the study explicitly connects predictive modeling to the concept of closed-loop DBS, where stimulation parameters could be dynamically adjusted based on inferred symptom state, providing a concrete example of how ML might move from assessment support to adaptive therapy optimization [30].

In synthesis, the provided literature depicts an ecosystem of ML/DL approaches targeting OCD diagnosis, severity estimation, remission prediction, and treatment optimization across modalities. Neuroimaging studies show that multi-index fMRI representations can improve discrimination in moderate samples, yet large-scale connectivity evidence cautions against overinterpreting group-level differences as actionable biomarkers. EEG work is rapidly maturing from heterogeneous, often non-interpretable pipelines toward externally validated, explainable frameworks. Biochemical and genetic studies extend the biomarker search into peripheral and

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

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

molecular domains, while clinical and cognitive predictors provide pragmatic routes to decision support using accessible routine measures. Across these directions, two cross-cutting imperatives emerge: (i) rigorous generalization testing across datasets and centers, and (ii) interpretable modeling that can justify predictions and guide clinical action rather than merely produce scores [20], [21], [22], [27]. These imperatives define the methodological agenda for translating computational advances into reliable, ethical, and clinically integrated OCD assessment systems.

3 Materials and Methods

3.1 Dataset Description

The dataset employed in this study is a comprehensive clinical repository specifically curated for the analysis of Obsessive-Compulsive Disorder (OCD). It comprises records from 1500 individuals who have been clinically diagnosed with OCD by qualified mental health professionals. The dataset is designed to capture a broad spectrum of demographic and clinical characteristics, thereby providing a rich and representative foundation for data-driven modeling and predictive analysis in psychiatric research. Its scale and diversity make it particularly suitable for developing and validating machine learning and deep learning models aimed at OCD-related classification tasks.

From a demographic perspective, the dataset includes several core attributes that reflect the socio-demographic diversity of the patient population. These features encompass patient age at the time of data collection, gender, ethnicity, marital status, and highest level of education attained. Such variables are known to influence both the manifestation of psychiatric symptoms and access to mental healthcare services, and their inclusion allows for a more nuanced understanding of OCD across different population subgroups. By incorporating these demographic dimensions, the dataset supports the exploration of potential associations between socio-demographic factors and OCD-related clinical patterns.

In addition to demographic information, the dataset contains detailed clinical and diagnostic features that characterize the patients' psychiatric profiles. These include the date of OCD diagnosis, which provides temporal context for disease onset and progression, as well as the duration of symptoms measured in months, offering insight into chronicity and illness trajectory. Furthermore, the dataset records any previous psychiatric diagnoses, enabling the identification of comorbid conditions that frequently coexist with OCD. The inclusion of family history of OCD serves to capture potential hereditary or environmental influences, which are widely recognized as important factors in the etiology and persistence of the disorder. Collectively, these clinical attributes form a multidimensional representation of each patient's diagnostic background and mental health history.

The prediction target in this study is an OCD-related classification outcome derived from the available clinical information. This outcome is formulated to support supervised learning and may represent a diagnostic category or a clinically meaningful classification related to OCD status or severity, depending on the modeling objective. Defining a clear and well-structured target variable is essential for enabling consistent model training and evaluation, as well as for ensuring clinical interpretability of the predictive outputs.

To facilitate robust model development and unbiased performance assessment, the dataset is partitioned into distinct subsets for training, validation, and testing. The training subset is used to learn model parameters and underlying data patterns, while the validation subset supports hyperparameter tuning and model selection. The testing subset is reserved exclusively for final performance evaluation, ensuring that reported outcomes reflect generalization to unseen data. This structured data splitting strategy is adopted to mitigate overfitting, enhance reproducibility, and provide a reliable estimate of model performance in real-world OCD classification scenarios.

Mental health disorders such as anxiety and depression represent a substantial and growing public health burden worldwide, affecting hundreds of millions of individuals across diverse demographic groups. Epidemiological evidence consistently demonstrates that these disorders are among the leading contributors to global disability, reduced quality of life, and increased healthcare utilization. Understanding diagnostic

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

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

patterns is therefore essential for informing healthcare planning, resource allocation, and early intervention strategies. In particular, comparing diagnosed and non-diagnosed populations provides insight into potential gaps in mental health recognition, access to care, and stigma-related barriers.

Figure 1 illustrates the distribution of patients with and without clinical diagnoses of anxiety and depression within the studied cohort. By presenting diagnosis status alongside patient counts, the figure offers a clear visual comparison of the relative prevalence of these two highly comorbid conditions. Such visualizations are valuable for contextualizing epidemiological findings, as anxiety and depression frequently coexist and share overlapping symptomatology, neurobiological mechanisms, and risk factors. Consequently, examining diagnostic rates side by side facilitates a more comprehensive interpretation of mental health trends and supports data-driven decision-making in both clinical and public health settings.

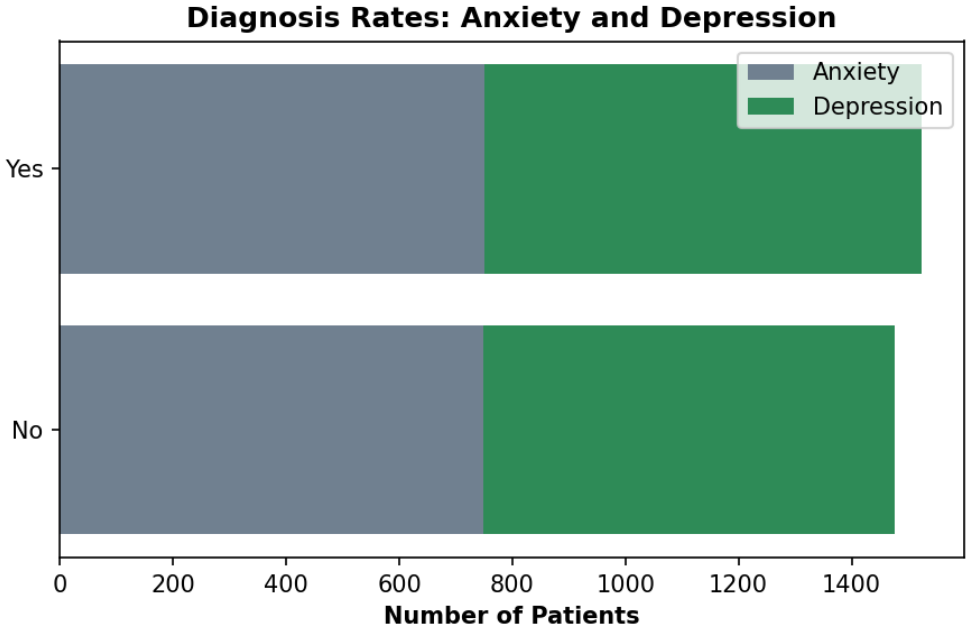

Figure 1: Distribution of patients with and without clinical diagnoses of anxiety and depression. The horizontal bar chart contrasts diagnosis status across the two conditions, highlighting differences in prevalence within the study cohort.

Symptom duration is a clinically important factor in obsessive–compulsive disorder (OCD), as prolonged illness course has been associated with greater symptom severity, functional impairment, and poorer treatment outcomes. The Yale–Brown Obsessive Compulsive Scale (Y-BOCS) is the most widely used clinician-administered instrument for quantifying OCD symptom severity, demonstrating strong reliability and validity across clinical and research settings. Examining the relationship between illness duration and Y-BOCS total scores can therefore provide valuable insights into the longitudinal progression of OCD and the extent to which chronicity relates to symptom burden.

Figure 2 presents a scatter plot illustrating the association between symptom duration (measured in months) and Y-BOCS total scores within the study sample. This visual representation enables an exploratory assessment of trends, variability, and potential non-linear patterns in symptom severity across different stages of illness. Such analyses are particularly relevant given prior evidence suggesting substantial heterogeneity in OCD trajectories, where long symptom duration does not uniformly correspond to higher severity, underscoring the influence of individual, neurobiological, and treatment-related factors.

Age is a fundamental demographic variable in clinical and epidemiological research, as it influences the onset, course, and clinical presentation of psychiatric disorders, as well as patterns of healthcare utilization. In mental health populations, age-related differences have been associated with variations in symptom severity, comorbidity profiles, and treatment response, underscoring the importance of characterizing age distributions when interpreting clinical findings. A clear understanding of the age composition of a study sample is therefore essential for assessing the generalizability and external validity of empirical results.

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

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

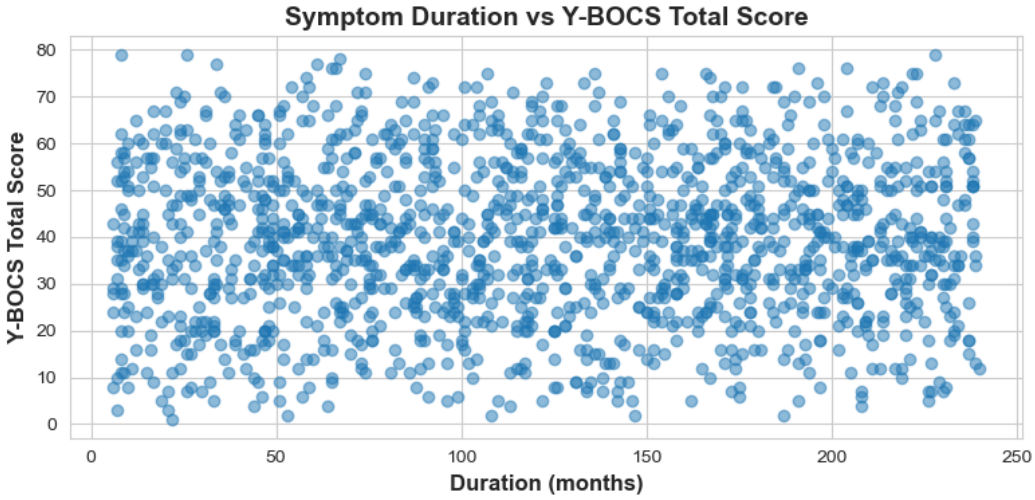

Figure 2: Scatter plot illustrating the relationship between symptom duration (in months) and Y-BOCS total scores. Each point represents an individual patient, highlighting the variability of symptom severity across different illness durations.

Figure 3 depicts the age distribution of patients included in the study using a histogram representation. This visualization provides an overview of the frequency of participants across age ranges, allowing for the identification of central tendencies, dispersion, and potential age-related sampling biases. Such descriptive analyses are particularly valuable in psychiatric research, where age of onset and illness trajectory often vary widely across individuals, potentially influencing both clinical outcomes and neurobiological correlates.

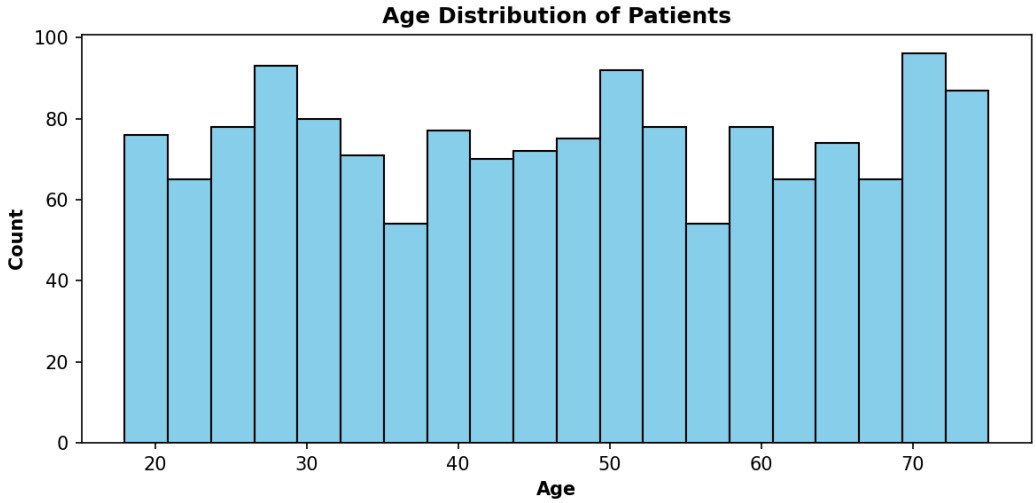

Figure 3: Histogram illustrating the age distribution of patients included in the study. The bars represent the frequency of individuals across age intervals, providing an overview of the demographic composition of the sample.

3.2 Data Preprocessing

Data preprocessing is a foundational component of the proposed machine learning pipeline, as the quality and structure of the input data directly influence the stability, convergence behavior, and predictive reliability of the developed models. In the context of psychiatric and clinical datasets, preprocessing assumes even greater importance due to the inherent variability in patient records, the presence of missing or incomplete information, and the coexistence of heterogeneous feature types. Accordingly, a comprehensive preprocessing framework

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

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

was designed to transform the raw OCD patient dataset into a form suitable for advanced machine learning and deep learning analysis, while preserving clinically meaningful information.

A primary preprocessing challenge arises from missing and incomplete records, which are common in real-world mental health datasets. Such missingness may stem from variations in clinical reporting practices, patient non-disclosure, or longitudinal gaps in medical documentation. Rather than discarding incomplete records—which could introduce bias and reduce the effective sample size—imputation strategies were employed to retain as much information as possible. Statistical imputation techniques were applied to numerical attributes by estimating missing values using distribution-based measures computed from the available data. These approaches aim to maintain the statistical properties of each feature while minimizing distortion of the underlying data structure. In parallel, model-based imputation methods were considered for cases where missingness exhibited more complex dependencies, allowing information from correlated variables to inform the imputation process. The combined use of statistical and model-driven imputation ensured robustness against different missing data patterns and enhanced dataset completeness without compromising data integrity.

The dataset also contains a substantial number of categorical variables, particularly within the demographic and clinical feature groups. Attributes such as gender, ethnicity, marital status, education level, previous psychiatric diagnoses, and family history of OCD are inherently categorical and cannot be directly processed by most machine learning and deep learning algorithms. To address this, categorical feature encoding techniques were applied to convert these variables into numerical representations suitable for model input. The selected encoding strategies were designed to preserve the semantic meaning of each category while avoiding the imposition of artificial ordinal relationships. This careful encoding process ensures that categorical distinctions contribute meaningfully to the learning process and that demographic and clinical heterogeneity is accurately reflected in the feature space.

Feature scaling and normalization were subsequently performed to ensure numerical stability and efficient optimization during model training. Psychiatric datasets often include numerical features with varying scales, such as age measured in years and symptom duration measured in months. Without proper scaling, features with larger numeric ranges may dominate gradient-based optimization processes, leading to biased learning dynamics and slower convergence. To mitigate this issue, numerical features were transformed into standardized or normalized forms, resulting in comparable value ranges across all inputs. This step is particularly critical for deep learning architectures, which are sensitive to input distributions and benefit from well-conditioned feature spaces. By promoting stable gradient updates and reducing numerical imbalance, scaling and normalization contribute to improved training efficiency and model robustness.

Beyond basic transformations, an initial correlation and redundancy analysis was conducted to examine relationships among input features prior to the application of metaheuristic feature selection algorithms. In high-dimensional clinical datasets, many variables may convey overlapping information, leading to multicollinearity and unnecessary model complexity. Such redundancy can obscure the contribution of truly informative features and negatively affect generalization performance. By analyzing feature correlations and dependencies, the preprocessing stage provided an early assessment of potential redundancy within the dataset. This analysis did not directly eliminate features but served as an informative precursor to the optimization phase, guiding the feature selection process toward more compact and discriminative subsets. Through this structured preprocessing pipeline, the dataset was transformed into a clean, stable, and informative representation, laying a robust foundation for subsequent modeling and optimization stages in OCD patient classification.

3.3 Deep Learning Models

To effectively model the complex relationships inherent in OCD-related demographic and clinical data, this study employs a carefully selected set of deep learning models that are well suited for capturing non-linear dependencies, hierarchical feature interactions, and potential temporal characteristics within structured datasets. The selection of these models is motivated by their established relevance in healthcare analytics and their complementary learning mechanisms, which collectively provide a robust basis for comparative evaluation. The four deep learning models considered in this work are DeepGBM, Artificial Neural Networks

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

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

(ANN), Convolutional Neural Networks (CNN), and Bidirectional Long Short-Term Memory networks (BiLSTM).

DeepGBM, a deep learning framework distilled by gradient boosting decision trees, constitutes the core modeling architecture investigated in this study. This model is specifically designed to address the limitations of conventional deep neural networks when applied to structured and tabular data, which are common in clinical and psychiatric research. DeepGBM leverages the strong feature interaction modeling capabilities of gradient boosting decision trees and transfers this knowledge into a neural network through a distillation process. As a result, the model benefits from both the expressive power of deep learning and the robustness, interpretability, and feature-awareness of tree-based methods. This hybrid design enables DeepGBM to effectively capture complex non-linear relationships among demographic and diagnostic variables while maintaining stable learning behavior, making it particularly suitable for OCD patient classification tasks where heterogeneous feature types coexist.

Artificial Neural Networks were included as a baseline deep learning model due to their general-purpose nature and extensive use in medical decision-support systems. ANNs consist of multiple layers of interconnected neurons that learn abstract representations of input data through weighted connections and non-linear activation functions. Their capacity to approximate complex decision boundaries allows them to model intricate relationships between patient attributes and diagnostic outcomes. In the context of this study, ANNs serve as a foundational reference for assessing the added value of more specialized architectures, offering insight into how increasing architectural complexity influences learning effectiveness in OCD-related classification problems.

Convolutional Neural Networks were incorporated to exploit their strength in automatic feature extraction and localized pattern learning. Although CNNs are most commonly associated with image and signal processing tasks, their convolutional operations can also be adapted to structured clinical data by learning local dependencies and feature groupings within the input space. By applying convolutional filters, CNNs are capable of identifying meaningful patterns across subsets of demographic and clinical features, potentially enhancing discriminative performance when such localized interactions exist. The inclusion of CNNs in this study allows for an examination of how spatially oriented feature learning mechanisms perform relative to fully connected and tree-distilled deep learning approaches in the psychiatric domain.

Bidirectional Long Short-Term Memory networks were selected to address the potential temporal and sequential aspects of the clinical data, such as diagnostic timelines and symptom duration trajectories. BiLSTM architectures extend traditional recurrent neural networks by incorporating memory cells and gating mechanisms that enable the retention and selective updating of information over long sequences, thereby alleviating vanishing gradient issues. The bidirectional structure allows the model to process sequences in both forward and backward directions, capturing contextual dependencies that may span across different temporal points. This capability is particularly relevant in mental health datasets, where past clinical events and future symptom progression can jointly inform diagnostic patterns and classification outcomes.

Among the evaluated deep learning models, DeepGBM was identified as the most appropriate candidate for subsequent optimization. Its architectural alignment with structured clinical data, combined with its ability to model complex feature interactions in a stable and interpretable manner, provides a strong foundation for enhancement through metaheuristic-based feature selection and hyperparameter tuning. Consequently, DeepGBM was selected as the primary model for optimization in later stages of this study, while the remaining models serve as comparative benchmarks for assessing the effectiveness of the proposed optimization framework.

3.4 Metaheuristic Optimization Algorithms

Metaheuristic optimization algorithms play a pivotal role in modern machine learning systems, particularly when addressing complex optimization problems characterized by high dimensionality, non-linearity, and the absence of explicit gradient information. In the context of deep learning-based OCD patient classification, these algorithms provide an effective mechanism for systematically exploring large hyperparameter search spaces and identifying configurations that yield robust and generalizable predictive models. Unlike

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

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

deterministic or exhaustive search strategies, metaheuristics are inherently stochastic and population-based, enabling them to escape local optima and adaptively balance exploration and exploitation throughout the optimization process.

3.4.1 Role of Metaheuristics in Hyperparameter Optimization

Hyperparameter optimization represents a critical challenge in the deployment of deep learning models for clinical decision support. Models such as DeepGBM, Artificial Neural Networks, Convolutional Neural Networks, and Bidirectional Long Short-Term Memory networks rely on a wide range of hyperparameters that directly influence learning dynamics, model capacity, and convergence behavior. These hyperparameters govern aspects such as learning rates, network depth, layer-wise neuron allocation, regularization strength, and optimization strategies. Improper hyperparameter selection can lead to unstable training, slow convergence, underfitting, or overfitting, ultimately limiting the clinical reliability of the resulting predictive system.

Traditional hyperparameter tuning methods, including manual tuning, grid search, and random search, are often inadequate for deep learning applications involving complex models and large datasets. These approaches either require extensive computational resources or fail to adequately explore the search space, increasing the likelihood of suboptimal solutions. In contrast, metaheuristic optimization algorithms frame hyperparameter tuning as a global optimization problem and employ adaptive search mechanisms that dynamically guide candidate solutions toward promising regions of the search space.

In this study, metaheuristic algorithms are utilized to automate the hyperparameter optimization process for DeepGBM as well as for the baseline deep learning models. By iteratively evaluating candidate hyperparameter configurations based on a predefined fitness function, these algorithms aim to identify parameter sets that enhance classification effectiveness while maintaining model stability. A central objective of this optimization strategy is to achieve a balance between classification performance and generalization capability. Rather than maximizing performance on training data alone, the optimization process emphasizes the discovery of hyperparameter configurations that support consistent behavior across unseen OCD patient records. Through their population-based search dynamics and probabilistic update mechanisms, metaheuristic algorithms are particularly well suited to navigating the complex and non-convex optimization landscapes associated with deep learning models in psychiatric applications.

3.5 Proposed NiOA–Based Optimization Framework

This study adopts the Ninja Optimization Algorithm (NiOA) as the core metaheuristic engine for hyperparameter optimization due to its strong theoretical foundation and its ability to dynamically balance exploration and exploitation in complex, high-dimensional search spaces. NiOA is a population-based metaheuristic inspired by the stealth, precision, and adaptability of Japanese ninjas, and it was explicitly designed to overcome two persistent challenges in metaheuristic optimization: premature convergence and entrapment in local optima. The mathematical formulation and workflow of NiOA used in this work strictly follow the original formulation presented in the NiOA reference paper :contentReference[oaicite:0]index=0.

Workflow of Hyperparameter Tuning

Within the proposed framework, NiOA is employed to automatically tune the hyperparameters of the DeepGBM model. Each ninja agent represents a candidate hyperparameter vector within a continuous search space defined by the allowable ranges of the DeepGBM hyperparameters. The optimization objective is formulated as a fitness function that reflects classification performance on the validation set, thereby guiding the search toward configurations that balance predictive accuracy and generalization.

The optimization process begins with the random initialization of a population of ninja agents. At each iteration, NiOA executes a sequence of exploration, mutation, and exploitation phases, followed by an adaptive solution update mechanism. These phases collectively enable both global search over unexplored regions of the hyperparameter space and local refinement around promising configurations. The best-performing solution is iteratively updated until a termination criterion, such as a maximum number of iterations or convergence threshold, is satisfied.

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

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

Exploration Phase

The exploration phase is designed to diversify the search and prevent premature convergence by encouraging agents to explore new and unvisited regions of the search space. The position update rule for a search agent during exploration is defined as:

$$L_s(t+1) = \begin{cases} L_s(t) + r_1 \cdot (L_s(t_1) - L_s(t_2)), & \text{if exploration condition is satisfied,} \\ \text{Random } L_s(t') \in FS_{met}, & \text{otherwise,} \end{cases} \tag{1}$$

where r_1 is a uniformly distributed random number in $[0, 1]$, t_1 and t_2 are randomly selected iteration indices, and FS_{met} denotes the feasible search space of the hyperparameters. This mechanism promotes large exploratory movements while maintaining feasibility constraints.

In parallel, an auxiliary variable is updated to further enhance exploration:

$$D_s(t+1) = D_s(t) + |D_s(t) + r_2 \cdot D_s(t)| \cdot \cos(2\pi t), \tag{2}$$

where $r_2 \in [0, 1]$ controls stochastic perturbations and the cosine term introduces oscillatory behavior that enables dynamic traversal of the search space.

Mutation Mechanism

To further increase population diversity and improve the ability to escape local optima, NiOA incorporates a mutation operator defined as:

$$N = \sum_{n=0}^{a} (-1)^n \frac{x \cdot (2n+1)}{2n+1}, \tag{3}$$

where a is a randomly selected integer parameter controlling mutation intensity. This operator introduces controlled perturbations that enhance exploration without destabilizing convergence.

Exploitation Phase

Once promising regions of the hyperparameter space have been identified, NiOA transitions to the exploitation phase, which focuses on refining high-quality solutions. The exploitation update rule is expressed as:

$$M_s(t+1) = J_1 M_s(t) + 2J_2 \cdot (M_s(t) + (M_s(t) + J_1)) \left(1 - \frac{M_s(t)}{M_s(t) + J_1}\right)^2, \tag{4}$$

where J_1 and J_2 are control parameters that regulate intensification strength. This formulation enables precise local search while preventing excessive exploitation that could lead to stagnation.

Adaptive Solution Update and Stagnation Handling

To maintain adaptability and avoid stagnation, the solution update mechanism is defined as:

$$R_s(t+1) = R_s(t) + (1 + R_s(t) + J_2) \cdot \exp(\cos(2\pi)), \tag{5}$$

introducing nonlinearity that allows dynamic adjustment to changes in the fitness landscape.

If no improvement is observed over consecutive iterations, NiOA applies an aggressive update strategy:

$$B_s(t+1) = L_s(t+1) + i \cdot n \cdot (L_s(t+1) - D_s(t+1)) + i \cdot n \cdot (M_s(t+1) + 2v_s \cdot R_s(t+1)), \tag{6}$$

where i, n, and v_s control the magnitude and direction of corrective movements. This mechanism significantly enhances the algorithm's ability to escape local optima.

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

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

Algorithm 1 Ninja Optimization Algorithm (NiOA)

Initialize parameters: population size N, maximum iterations T, random initial positions $L_{s,t}$, $D_{s,t}$, random factors r_1, r_2, control parameters J_1, J_2, stagnation factor n, mutation parameter a, velocity factor v_s, and best solution B_s
Set iteration counter $t = 0$
while $t < T$ **do**
  Exploration Phase:
  for each agent $s = 1, 2, \ldots, N$ **do**
    Update position $L_{s,t+1}$:
    $L_{s,t+1} = L_{s,t} + r_1 \cdot (L_{s,t_1} - L_{s,t_2})$ **or** randomly generate $L'_s \in F_{met_s}$
    Update position $D_{s,t+1}$:
    $D_{s,t+1} = D_{s,t} + |D_{s,t} + r_2 \cdot D_{s,t}| \cdot \cos(2\pi t)$
  end for
  Mutation Phase:
  Perform mutation using:
  $N = \sum_{n=0}^{a-1} (-1)^n \cdot \frac{2n+1}{2n+1}$
  Exploitation Phase:
  Update $M_{s,t+1}$:
  $M_{s,t+1} = J_1 \cdot M_{s,t} + 2J_2 \cdot M_{s,t} + M_{s,t} + J_1 \cdot (1 - M_{s,t}) \cdot M_{s,t} + J_1^2$
  Resource Update:
  Update $R_{s,t+1}$:
  $R_{s,t+1} = R_{s,t} + 1 + R_{s,t} + J_2 \cdot \exp(\cos(2\pi))$
  Best Solution Update:
  if no improvement is observed for three consecutive iterations **then**
    Update best solution $B_{s,t+1}$:
    $B_{s,t+1} = L_{s,t+1} + i \cdot n \cdot (L_{s,t+1} - D_{s,t+1}) + i \cdot n \cdot M_{s,t+1} + 2v_s \cdot R_{s,t+1}$
  end if
  $t \leftarrow t + 1$
end while
Return the best solution B_s

Compared with traditional swarm-based optimizers such as PSO and BA, NiOA introduces a more structured and adaptive transition between exploration and exploitation phases. Unlike PSO, which relies heavily on velocity updates guided by global and personal best positions, NiOA employs oscillatory motion, mutation, and stagnation-aware updates to enhance search robustness. In contrast to BA, which adapts search behavior through echolocation-inspired parameters, NiOA explicitly models precision-driven exploitation and stealth-based exploration using mathematically distinct operators. These characteristics make NiOA particularly well suited for hyperparameter optimization in deep learning models, where search landscapes are highly non-convex and sensitive to parameter interactions.

Overall, the proposed NiOA-based optimization framework provides a mathematically grounded, adaptive, and robust mechanism for tuning DeepGBM hyperparameters, forming a critical component of the proposed OCD classification pipeline.

3.5.1 State-of-the-Art Metaheuristic Algorithms Used

To construct a robust and comprehensive hyperparameter optimization framework, this study incorporates a selection of state-of-the-art metaheuristic algorithms that have demonstrated effectiveness across a wide range of machine learning and optimization tasks. The chosen algorithms—Multiverse Optimization, Bat Algorithm, and Particle Swarm Optimization—are grounded in distinct natural and physical metaphors, offering complementary search behaviors and exploration–exploitation trade-offs.

Multiverse Optimization (MVO) is a population-based metaheuristic inspired by concepts from cosmology, particularly theories related to multiple universes and their interactions. In MVO, each candidate solution is conceptualized as a universe, and the optimization process is governed by mechanisms analogous to

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

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

white holes, black holes, and wormholes. White hole and black hole operators facilitate information exchange between universes based on their relative fitness, promoting the transfer of desirable traits from high-quality solutions to weaker ones. Wormhole mechanisms, on the other hand, enable local search around the best-performing universes, enhancing exploitation while preserving global exploration. This structured balance allows MVO to effectively navigate complex hyperparameter spaces and reduce the risk of premature convergence.

The Bat Algorithm (BA) is a bio-inspired metaheuristic that simulates the echolocation behavior of bats during foraging. In this algorithm, each bat represents a candidate solution that moves through the search space by dynamically adjusting its position and velocity in response to frequency, loudness, and pulse emission rates. These parameters are adaptively modified as the search progresses, allowing the algorithm to transition from broad global exploration to focused local exploitation. The Bat Algorithm's ability to self-regulate its search behavior makes it particularly suitable for hyperparameter optimization problems where the optimal solution lies within a highly irregular and multimodal landscape.

Particle Swarm Optimization (PSO) is a well-established metaheuristic inspired by the collective movement of social organisms such as bird flocks and fish schools. In PSO, candidate solutions, referred to as particles, traverse the search space by updating their velocities based on both individual experience and shared information from the swarm. Each particle is influenced by its own best-known position as well as the globally best position identified by the swarm, fostering cooperative learning and efficient convergence. PSO's simplicity, scalability, and strong convergence properties have led to its widespread adoption for hyperparameter optimization in machine learning and deep learning applications.

By integrating Multiverse Optimization, Bat Algorithm, and Particle Swarm Optimization into the hyperparameter tuning framework, this study establishes a diverse optimization environment capable of addressing the complexities inherent in deep learning-based OCD classification. The use of multiple metaheuristic strategies enables a systematic examination of different search behaviors and provides a robust foundation for enhancing model generalization and stability in subsequent stages of the proposed framework.

3.6 Evaluation Metrics

A rigorous and comprehensive evaluation framework is essential for assessing the effectiveness and clinical reliability of machine learning and deep learning models, particularly in the context of mental health applications such as OCD patient classification. Given the sensitive nature of psychiatric decision-making, reliance on a single performance indicator may lead to misleading conclusions regarding model suitability. Therefore, this study adopts a set of complementary evaluation metrics derived from the confusion matrix to ensure a balanced and interpretable assessment of predictive behavior.

Let TP denote the number of true positive instances, corresponding to OCD cases correctly identified by the model, while TN represents true negative instances that are correctly classified as non-target cases. Similarly, FP refers to false positive instances in which the model incorrectly predicts a positive outcome, and FN denotes false negative instances where actual positive cases are missed by the model. These four fundamental quantities form the basis for computing all evaluation metrics used in this study.

Accuracy is employed as a global performance indicator that measures the proportion of correctly classified instances among all predictions. While accuracy provides an intuitive overview of model correctness, it may be insufficient on its own, particularly in clinical datasets where class imbalance or asymmetric misclassification costs are common. For this reason, additional metrics are incorporated to capture class-specific performance characteristics.

Sensitivity, also known as the true positive rate (TPR), quantifies the model's ability to correctly identify positive cases. In the context of OCD classification, high sensitivity is crucial, as failing to detect affected individuals may result in delayed diagnosis and inadequate treatment. Complementary to sensitivity, specificity or the true negative rate (TNR) measures the model's ability to correctly identify negative cases. High specificity is important for minimizing false alarms and reducing the risk of unnecessary clinical interventions.

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

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

Positive Predictive Value (PPV) evaluates the proportion of correctly predicted positive cases among all positive predictions made by the model. This metric reflects the reliability of positive classifications and is particularly relevant when assessing the trustworthiness of automated decision-support systems. Conversely, Negative Predictive Value (NPV) measures the proportion of correctly predicted negative cases among all negative predictions, indicating the confidence that can be placed in negative outcomes produced by the model.

Finally, the F-score is included as a balanced performance metric that combines sensitivity and precision into a single harmonic measure. By simultaneously accounting for false positives and false negatives, the F-score provides a more comprehensive representation of classification effectiveness, especially in scenarios where both types of errors carry significant clinical implications.

The mathematical formulations of all evaluation metrics employed in this study are presented in Table 1. This table serves as a formal reference for the quantitative assessment framework adopted throughout the experimental analysis.

Table 1: Classification evaluation metrics and their mathematical definitions

Metric	**Mathematical Equation**
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
Sensitivity (TPR)	$\frac{TP}{TP+FN}$
Specificity (TNR)	$\frac{TN}{TN+FP}$
Positive Predictive Value (PPV)	$\frac{TP}{TP+FP}$
Negative Predictive Value (NPV)	$\frac{TN}{TN+FN}$
F-Score	$\frac{2TP}{2TP+FP+FN}$

The use of these evaluation metrics provides a multidimensional perspective on model performance, enabling a nuanced analysis of strengths and limitations across different classification aspects. By jointly considering overall accuracy, class-specific detection capability, and predictive reliability, the evaluation framework ensures that the assessed models are not only statistically effective but also clinically meaningful. This comprehensive metric selection establishes a solid foundation for comparing baseline and optimized deep learning models in the subsequent experimental analysis.

4 Experimental Results

4.1 Baseline Model Performance (Before Optimization)

This subsection provides an in-depth quantitative analysis of the baseline performance of the evaluated deep learning models prior to the application of any metaheuristic-based feature selection or hyperparameter optimization strategies. Establishing a strong and transparent baseline is essential for accurately assessing the impact of the proposed optimization framework and for identifying the specific weaknesses that necessitate further methodological enhancement.

Four deep learning models—Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), Bidirectional Long Short-Term Memory networks (BiLSTM), and DeepGBM—were evaluated under identical experimental settings using the full feature set and non-optimized hyperparameters. Performance evaluation was conducted using six complementary metrics, namely Accuracy, Sensitivity (TPR), Specificity (TNR), Positive Predictive Value (PPV), Negative Predictive Value (NPV), and F-Score, ensuring a comprehensive assessment of classification behavior.

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

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

The baseline quantitative results are summarized in Table 2, which serves as the primary reference for subsequent comparative analyses.

Table 2: Baseline performance of deep learning models before optimization

Model	Accuracy	Sensitivity (TPR)	Specificity (TNR)	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	0.876577840	0.867052023	0.885558583	0.877192982	0.876010782	0.872093023
ANN	0.859491779	0.848765432	0.869565217	0.859375000	0.859598854	0.854037267

A detailed examination of the numerical results reveals that DeepGBM consistently outperforms the remaining baseline models across all evaluation metrics. Specifically, DeepGBM achieves the highest classification accuracy of 0.897009967, indicating that nearly 89.7% of OCD-related instances are correctly classified. Its sensitivity value of 0.890410959 demonstrates a strong capability to correctly identify positive cases, which is critical in psychiatric applications where missed diagnoses can have serious clinical consequences. At the same time, DeepGBM attains the highest specificity of 0.903225806, reflecting its effectiveness in correctly recognizing negative cases and reducing false-positive classifications. The high PPV (0.896551724) and NPV (0.897435897) further indicate that both positive and negative predictions generated by DeepGBM are reliable. This balanced behavior is consolidated by an F-score of 0.893470790, confirming a strong trade-off between sensitivity and precision.

The BiLSTM model demonstrates the second-best baseline performance, achieving an accuracy of 0.885135135 and an F-score of 0.881118881. Its sensitivity of 0.875000000 and specificity of 0.894736842 indicate a reasonably balanced detection capability for both positive and negative cases. While these values are competitive, they remain consistently lower than those of DeepGBM, suggesting that although bidirectional temporal modeling contributes positively, it does not fully capture the structured feature interactions present in the OCD dataset.

The CNN model achieves an accuracy of 0.876577840 and an F-score of 0.872093023, placing it below BiLSTM but above ANN. Its sensitivity (0.867052023) and specificity (0.885558583) indicate moderate classification effectiveness. The comparatively lower PPV (0.877192982) and NPV (0.876010782) suggest that while CNNs can learn localized feature patterns, their inductive bias may not be optimally aligned with purely tabular demographic and clinical data, limiting their discriminative power in this context.

The ANN model records the weakest baseline performance across all metrics, with an accuracy of 0.859491779 and an F-score of 0.854037267. Its sensitivity (0.848765432) and specificity (0.869565217) are noticeably lower than those of the other models, indicating a reduced ability to accurately distinguish between OCD-related classes. These results highlight the limitations of fully connected architectures when applied to high-dimensional clinical datasets without specialized mechanisms for feature interaction modeling or temporal representation.

Although the baseline results—particularly those of DeepGBM—demonstrate reasonably strong predictive capability, several limitations are evident from the numerical analysis. First, the performance gap between models indicates sensitivity to architectural design and hyperparameter configurations, suggesting that the baseline settings may not be optimal. Second, none of the evaluated models incorporate explicit mechanisms for feature selection, leaving them susceptible to feature redundancy and noise, which may constrain further performance improvements. Finally, the observed metric values indicate room for enhancement in both sensitivity and specificity, motivating the integration of metaheuristic optimization techniques aimed at refining hyperparameters and reducing feature dimensionality. These observations collectively justify the transition to the optimized modeling framework presented in the subsequent sections.

Comprehensive evaluation of machine learning models requires the simultaneous consideration of multiple performance metrics, as reliance on a single indicator such as accuracy may obscure clinically relevant strengths or weaknesses. Metrics including sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and F-score are particularly important in medical and psychiatric classification tasks, where class imbalance and asymmetric error costs are common. Visual comparison of these metrics across

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

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

different models facilitates a more nuanced interpretation of classifier behavior and supports transparent model selection.

Figure 4 presents a detailed heatmap summarizing the performance of four predictive models—DeepGBM, BiLSTM, CNN, and ANN—across six evaluation metrics. The color intensity and numerical annotations enable direct comparison of model performance patterns, highlighting relative strengths in discrimination, predictive reliability, and overall balance. Such heatmap-based representations have been widely adopted in machine learning research due to their effectiveness in conveying high-dimensional evaluation results in a compact and interpretable form.

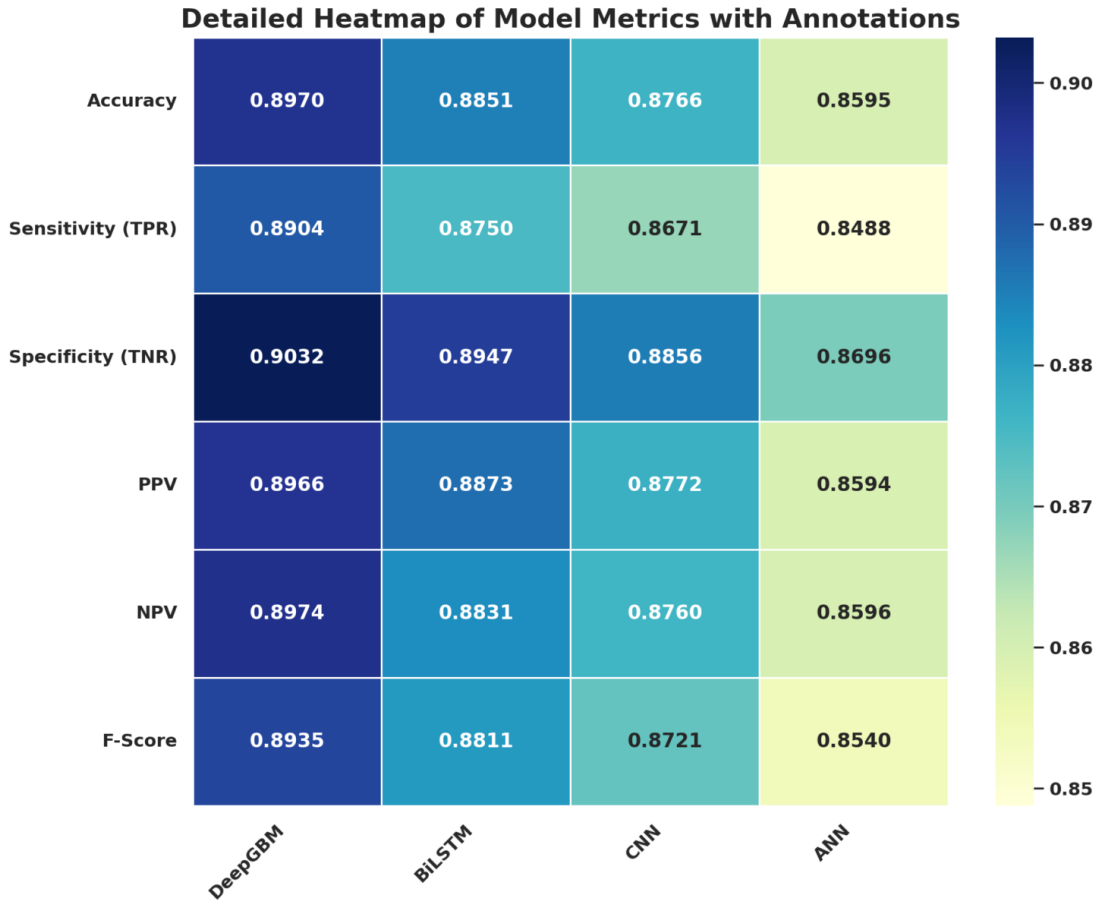

Figure 4: Heatmap visualization of performance metrics for four classification models (DeepGBM, BiLSTM, CNN, and ANN). Rows correspond to evaluation metrics, including accuracy, sensitivity (TPR), specificity (TNR), PPV, NPV, and F-score, while columns represent the evaluated models. Numerical annotations indicate exact metric values.

Understanding the interrelationships among evaluation metrics is essential for the robust interpretation of machine learning model performance, particularly in medical and psychiatric classification tasks. Metrics such as accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and F-score are mathematically and conceptually interdependent, and their pairwise associations can reveal redundancy, trade-offs, or consistency in classifier behavior. Analyzing these relationships provides deeper insight beyond isolated metric values and supports more informed model comparison and validation.

Figure 5 illustrates a pairwise comparison of the evaluated performance metrics using scatter plots with overlaid regression lines and marginal distributions. This visualization facilitates the assessment of linear associations, dispersion patterns, and metric stability across models. Pairplot-based analyses are widely employed in exploratory data analysis due to their effectiveness in revealing structural relationships in multivariate performance data, thereby enhancing transparency and interpretability in machine learning evaluation workflows.

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

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

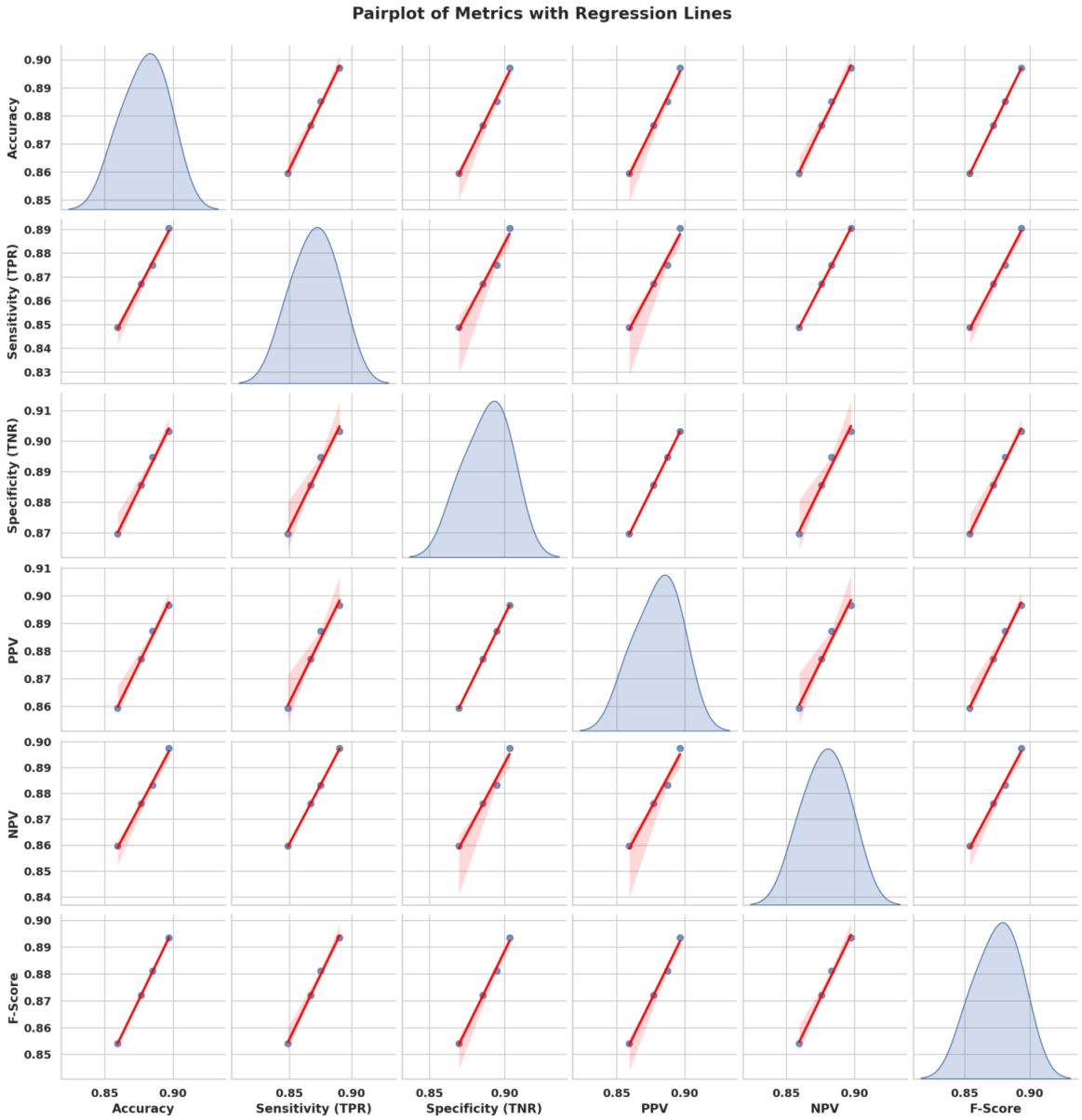

Figure 5: Pairplot visualization of classification performance metrics, including accuracy, sensitivity (TPR), specificity (TNR), PPV, NPV, and F-score. Off-diagonal panels display pairwise scatter plots with fitted regression lines, while diagonal panels show the marginal distributions of each metric.

Beyond point estimates of performance metrics, analyzing their empirical distributions provides deeper insight into model stability, variability, and robustness. Kernel density estimation (KDE) is a non-parametric technique widely used to approximate the underlying probability density function of continuous variables, offering a smooth and informative representation of metric dispersion. In machine learning evaluation, KDE plots enable comparative assessment of multiple performance measures by highlighting central tendencies, overlap, and spread across models.

Figure 6 presents KDE curves for six commonly used classification metrics, including accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and F-score. By visualizing these metrics within a unified density-based framework, the figure facilitates a nuanced comparison of their distributions and relative consistency. Such distributional analyses are particularly valuable in medical and psychiatric classification tasks, where small variations in performance metrics can have significant clinical implications.

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

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

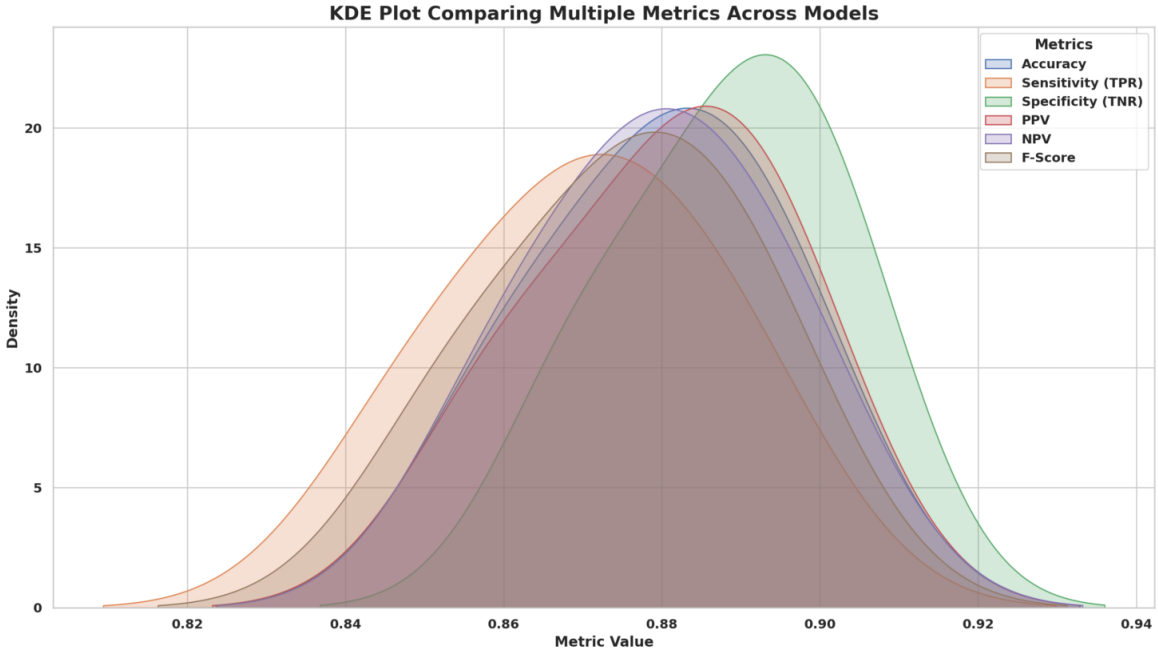

Figure 6: Kernel density estimation (KDE) plots comparing the distributions of classification performance metrics across models. Each curve represents the estimated probability density of a specific metric, enabling visual assessment of central tendency, dispersion, and overlap among metrics.

4.2 Optimized Model Analysis

This subsection presents a detailed analysis of the optimized DeepGBM models obtained through the integration of different metaheuristic optimization algorithms. The primary focus is on evaluating the effectiveness of the proposed NiOA–DeepGBM framework in comparison with other optimized variants, namely MVO+DeepGBM, BA+DeepGBM, and PSO+DeepGBM. The analysis emphasizes both quantitative performance improvements and their clinical relevance, with particular attention to robustness, reliability, and potential applicability in real-world OCD assessment scenarios.

All optimized models were evaluated using the same experimental protocol and performance metrics described in Section 2.7 to ensure a fair comparison. The numerical results of the optimized models are reported in Table 3.

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

Model	Accuracy	Sensitivity (TPR)	Specificity (TNR)	PPV	NPV	F-Score
NiOA + DeepGBM	0.977851463	0.976301736	0.979294908	0.977737665	0.977957156	0.97701917
MVO + DeepGBM	0.967023521	0.962031811	0.971223022	0.965665236	0.968158348	0.96384509
BA + DeepGBM	0.953973713	0.944088954	0.962031811	0.952986023	0.954765340	0.94851662
PSO + DeepGBM	0.944336604	0.932173295	0.954018656	0.941647734	0.946438587	0.93688656

A careful examination of the results in Table 3 clearly demonstrates the superiority of the NiOA–DeepGBM framework over all other optimized variants. NiOA+DeepGBM achieves the highest classification accuracy of 0.977851463, indicating that approximately 97.8% of OCD-related instances are correctly classified. This represents a substantial improvement over all competing optimization strategies and reflects the effectiveness of NiOA in identifying optimal hyperparameter configurations for DeepGBM. The sensitivity of 0.976301736 further highlights the model's strong ability to correctly detect positive OCD cases, which is of paramount importance in clinical settings where false negatives can lead to delayed diagnosis and inadequate treatment.

In parallel, NiOA+DeepGBM records the highest specificity value of 0.979294908, demonstrating an exceptional capacity to correctly identify negative cases and minimize false-positive outcomes. This balanced

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

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

behavior is further reinforced by the high PPV of 0.977737665 and NPV of 0.977957156, indicating that both positive and negative predictions generated by the model are highly reliable. The resulting F-score of 0.977019173 confirms a near-optimal balance between sensitivity and precision, underscoring the robustness and consistency of the NiOA-optimized model.

The MVO+DeepGBM model ranks second in overall performance, achieving an accuracy of 0.967023521 and an F-score of 0.963845099. Although these values indicate strong predictive capability, they remain noticeably lower than those achieved by NiOA+DeepGBM. Specifically, the sensitivity (0.962031811) and specificity (0.971223022) suggest effective classification, yet the marginally reduced values imply that MVO is less capable of fine-grained hyperparameter tuning compared to NiOA, particularly in highly non-linear optimization landscapes.

The BA+DeepGBM model exhibits moderate performance improvements relative to the baseline but falls behind both NiOA and MVO optimization strategies. With an accuracy of 0.953973713 and an F-score of 0.948516625, BA+DeepGBM demonstrates reasonable classification capability; however, its lower sensitivity (0.944088954) indicates a comparatively higher risk of misclassifying positive OCD cases. This limitation may reduce its suitability for clinical applications where high recall is critical.

Similarly, PSO+DeepGBM yields the lowest performance among the optimized models, with an accuracy of 0.944336604 and an F-score of 0.936886562. While PSO improves DeepGBM performance compared to the non-optimized baseline, its sensitivity (0.932173295) and specificity (0.954018656) suggest weaker generalization and less stable convergence when compared with the other metaheuristic approaches. This behavior can be attributed to PSO's reliance on velocity-based updates, which may struggle to balance exploration and exploitation in highly complex hyperparameter spaces.

From a clinical perspective, the results highlight the practical relevance of the NiOA+DeepGBM framework for OCD assessment. The consistently high sensitivity and specificity values indicate that the model can reliably support both the identification of affected individuals and the exclusion of non-affected cases, thereby reducing diagnostic uncertainty. Moreover, the robustness reflected in uniformly high PPV and NPV values suggests that the model's predictions can be trusted across diverse patient profiles. Collectively, these characteristics position NiOA+DeepGBM as a highly reliable and clinically meaningful decision-support tool, capable of enhancing diagnostic accuracy and supporting personalized mental healthcare interventions.

Overall, the optimized model analysis confirms that the proposed NiOA-based optimization framework provides superior performance and robustness compared to other state-of-the-art metaheuristic optimization strategies, justifying its selection as the primary optimization approach in this study.

A rigorous comparison of predictive models requires evaluation across multiple complementary performance metrics, particularly in high-stakes classification tasks such as medical diagnosis. While accuracy provides an overall measure of correctness, metrics such as sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and F-score offer more granular insight into error characteristics and class-wise performance. This multidimensional evaluation is essential for assessing model reliability, robustness, and clinical applicability, especially in the presence of class imbalance.

Figure 7 summarizes the comparative performance of four DeepGBM-based hybrid models across six evaluation metrics. By presenting each metric in a dedicated subplot, the figure enables direct visual comparison of model behavior, highlighting consistent performance patterns as well as metric-specific trade-offs. Such bar-chart-based representations are commonly employed in machine learning studies to facilitate transparent and interpretable benchmarking of competing models.

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

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

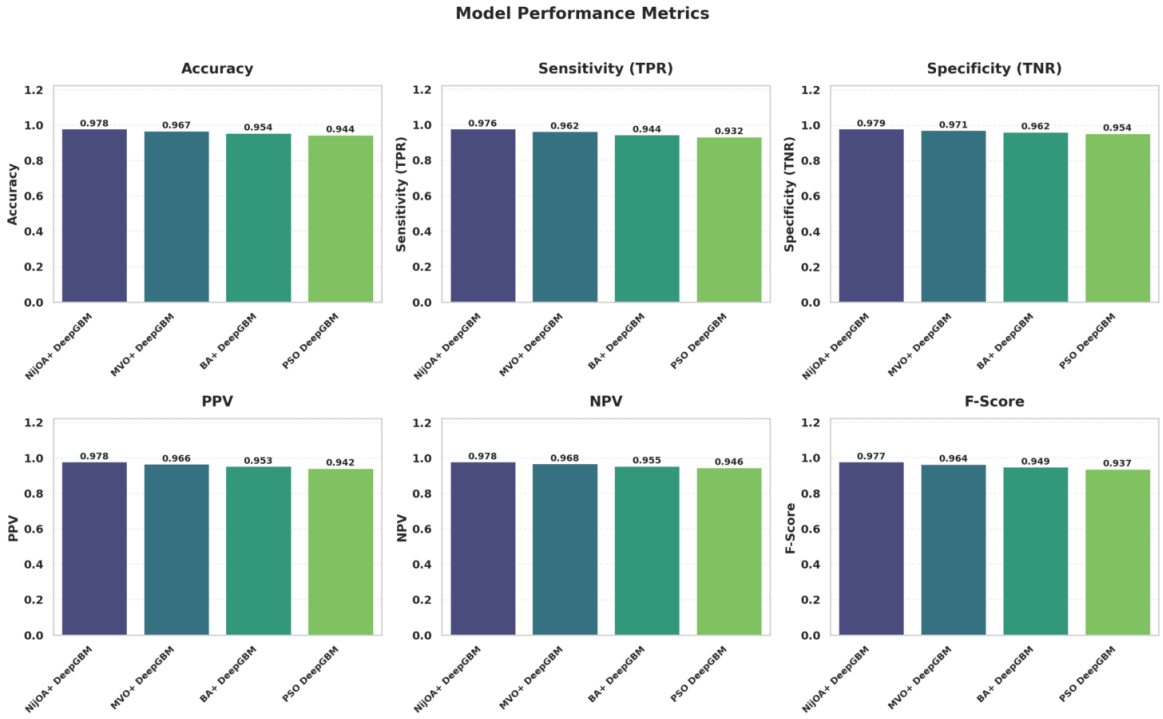

Figure 7: Bar chart comparison of classification performance metrics for four hybrid DeepGBM models. Subplots display accuracy, sensitivity (TPR), specificity (TNR), positive predictive value (PPV), negative predictive value (NPV), and F-score, with numerical annotations indicating exact metric values.

Evaluating the relationships among performance metrics is critical for understanding redundancy, complementarity, and structural similarity in model assessment. In classification problems, many commonly used metrics—such as accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and F-score—are mathematically interrelated, which can lead to strong correlations and overlapping informational content. Hierarchical clustering provides a principled, data-driven approach to uncovering these relationships by grouping metrics based on similarity, thereby offering insights into metric dependence and interpretability.

Figure 8 illustrates a correlation heatmap of model performance metrics augmented with hierarchical clustering dendrograms. The combined visualization enables simultaneous examination of pairwise metric correlations and their higher-level grouping structure. Such heatmap–dendrogram representations are widely used in machine learning and biomedical research to identify clusters of related variables, reduce dimensional redundancy, and support more informed metric selection in model evaluation pipelines.

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

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

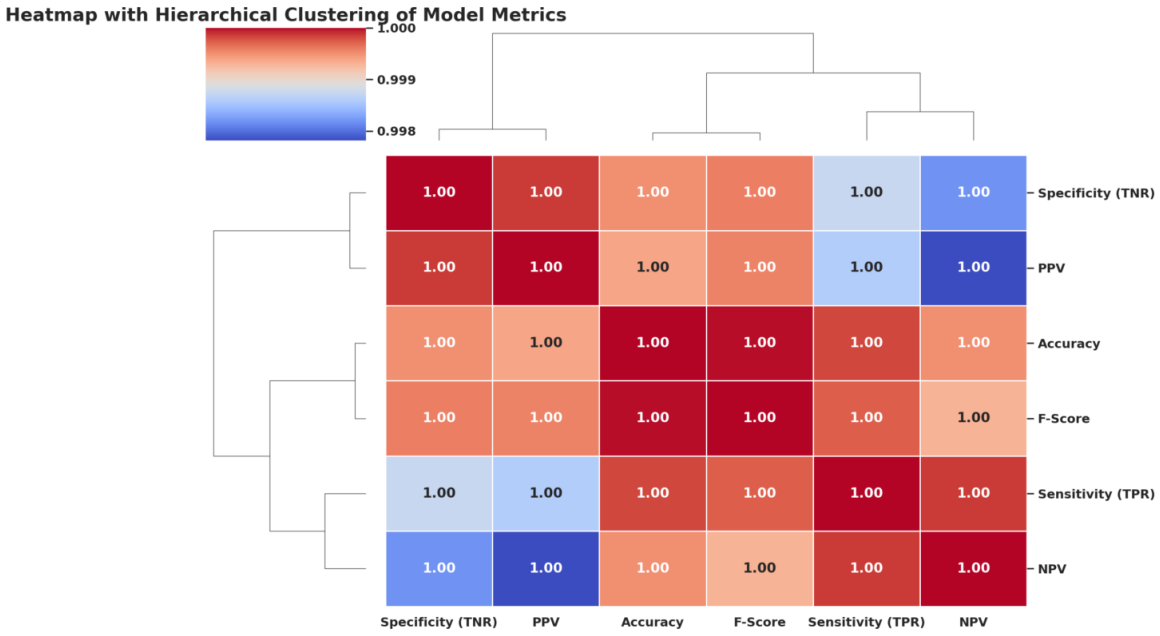

Figure 8: Correlation heatmap of classification performance metrics with hierarchical clustering. Cell values represent pairwise correlation coefficients, while dendrograms indicate metric similarity based on hierarchical linkage, revealing clusters of closely related evaluation measures.

Exploring the joint distributions of performance metrics provides valuable insight into the internal consistency and dependency structure of model evaluation outcomes. In classification tasks, metrics such as accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and F-score are inherently interrelated, as they are derived from overlapping components of the confusion matrix. Pairwise kernel density estimation (KDE) plots offer an effective exploratory tool for visualizing these relationships by simultaneously representing marginal distributions and bivariate associations.

Figure 9 presents a pairwise KDE visualization of the evaluated model performance metrics. The diagonal panels depict the univariate density estimates of individual metrics, while the off-diagonal panels illustrate pairwise relationships, highlighting the degree of linear association and distributional overlap among metrics. Such visual analyses are widely used in machine learning and biomedical research to assess metric redundancy, detect anomalies, and support transparent interpretation of multi-metric evaluation results.

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

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

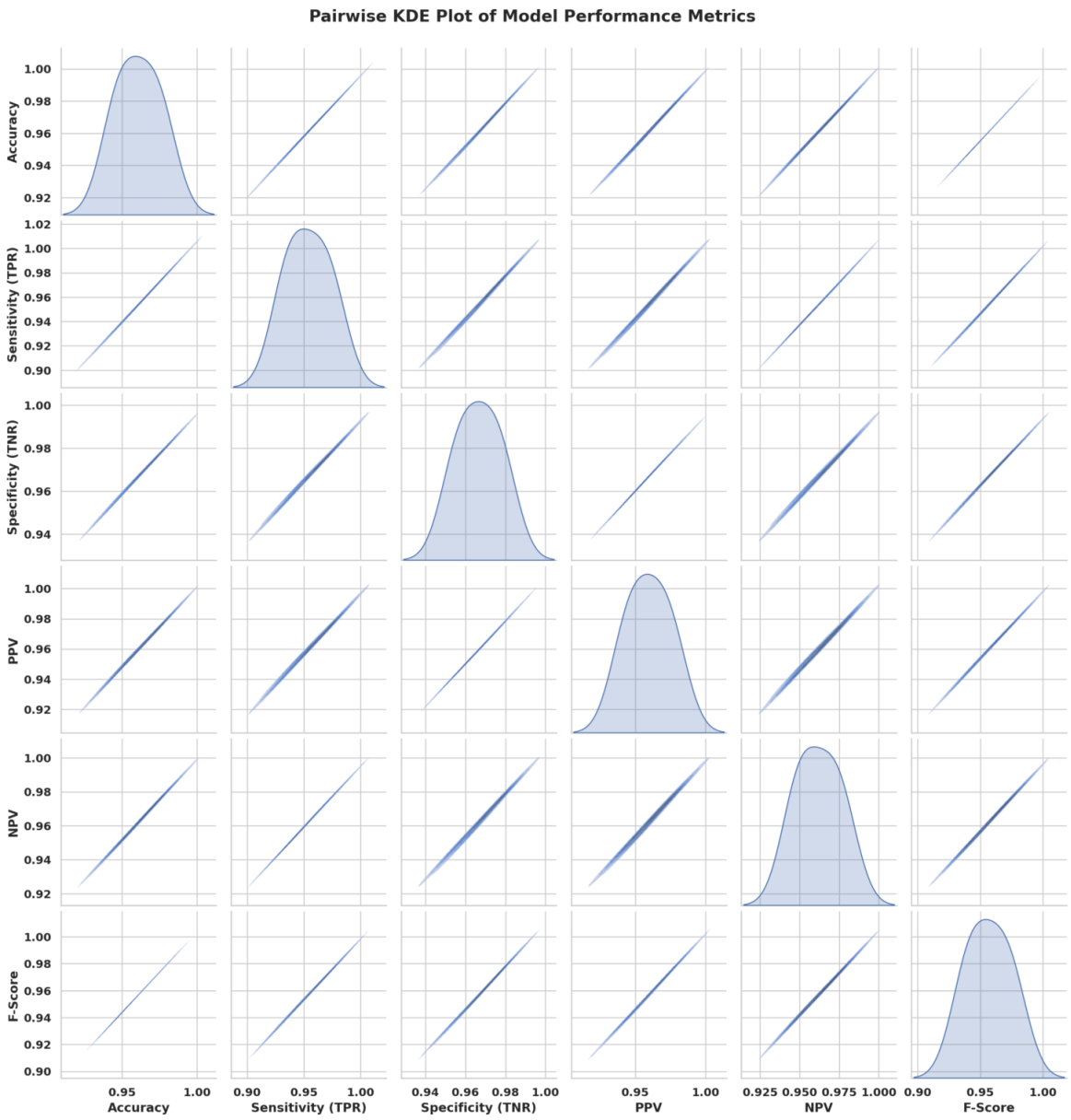

Figure 9: Pairwise kernel density estimation (KDE) plot of classification performance metrics. Diagonal panels show marginal density distributions for each metric, while off-diagonal panels represent pairwise relationships, enabling assessment of dependency and consistency across evaluation measures.

5 Discussion

This section provides a comprehensive interpretation of the experimental findings by examining the differences between baseline and optimized models, discussing the clinical relevance of the observed performance improvements, situating the results within the broader landscape of psychiatric machine learning research, and outlining the main limitations and methodological considerations of the proposed framework. The comparative analysis between baseline and optimized models clearly demonstrates the critical role of metaheuristic-driven hyperparameter optimization in enhancing deep learning performance for OCD-related classification. At the baseline stage, DeepGBM emerged as the strongest model among all evaluated architectures, achieving an accuracy of 0.8970 and an F-score of 0.8935. Although these results indicate a solid predictive capability, noticeable gaps remained in sensitivity and specificity, reflecting the inherent limitations of relying on non-optimized hyperparameter configurations and the full, unfiltered feature space.

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

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

The introduction of metaheuristic optimization strategies led to consistent and substantial improvements across all performance metrics. Among the optimized models, the Ninja Optimization Algorithm–optimized DeepGBM achieved the highest performance, with accuracy increasing from 0.8970 to 0.9779 and the F-score improving from 0.8935 to 0.9770. Sensitivity and specificity also exhibited marked gains, reaching 0.9763 and 0.9793, respectively. These improvements indicate that the optimization process effectively refined the model's decision boundaries, enabling more precise discrimination between OCD-related classes while maintaining strong generalization. In contrast, alternative metaheuristic optimizers such as Multiverse Optimization, Bat Algorithm, and Particle Swarm Optimization also enhanced DeepGBM performance but to a lesser extent. For instance, Multiverse Optimization–optimized DeepGBM achieved an accuracy of 0.9670, while Bat Algorithm– and Particle Swarm Optimization–based variants attained accuracies of 0.9540 and 0.9443, respectively. The consistent performance hierarchy across all metrics suggests that the superior exploration–exploitation balance of the Ninja Optimization Algorithm enables it to more effectively navigate the complex, non-convex hyperparameter space of DeepGBM.

From a clinical perspective, the observed performance gains are highly significant. In psychiatric applications, both false negatives and false positives carry substantial consequences. A false negative may delay diagnosis and treatment, potentially worsening symptom severity and long-term outcomes, while a false positive may lead to unnecessary psychological burden and inappropriate intervention. The optimized NiOA–DeepGBM model achieves simultaneously high sensitivity (0.9763) and specificity (0.9793), indicating a strong capacity to correctly identify affected individuals while minimizing misclassification of non-affected cases. The high Positive Predictive Value (0.9777) and Negative Predictive Value (0.9780) further enhance the clinical reliability of the proposed framework, suggesting that both positive and negative predictions can be trusted with a high degree of confidence. These characteristics position the optimized model as a valuable decision-support tool that can complement clinician expertise rather than replace it. By providing consistent and objective predictions derived from demographic and clinical data, the proposed system may help reduce diagnostic variability, support early intervention, and contribute to more personalized mental healthcare strategies for individuals with OCD.

When compared with existing studies in psychiatric machine learning, the results obtained in this work are highly competitive and, in many cases, superior. Previous research has commonly employed traditional machine learning classifiers or standalone deep learning models, often reporting moderate accuracy and limited generalization due to small sample sizes, feature redundancy, or insufficient hyperparameter tuning. While recent studies have begun to explore metaheuristic optimization for psychiatric prediction tasks, many focus on single optimization objectives or simpler model architectures.

The proposed framework distinguishes itself by combining a deep learning model specifically designed for structured data with a powerful and adaptive metaheuristic optimizer. The achieved accuracy of 0.9779 and F-score of 0.9770 exceed the performance levels typically reported for OCD-related classification tasks, highlighting the effectiveness of integrating advanced optimization strategies into psychiatric modeling pipelines. Moreover, the consistent improvement over alternative metaheuristic approaches underscores the importance of optimizer design and search dynamics in achieving robust and clinically meaningful performance.

Despite the promising results, several limitations should be acknowledged. First, the study relies on a single dataset comprising 1500 clinically diagnosed OCD patients. Although this dataset is relatively large for psychiatric research, external validation on independent cohorts is necessary to confirm the generalizability of the proposed framework across different populations, healthcare settings, and demographic distributions. Second, the current study focuses on a binary classification task. While this formulation is suitable for initial diagnostic support, OCD is a heterogeneous disorder with varying symptom dimensions and severity levels. Extending the framework to multi-class or ordinal severity prediction would provide more granular clinical insights but may introduce additional modeling and optimization challenges.

Finally, although the Ninja Optimization Algorithm demonstrated superior performance, metaheuristic optimization inherently involves stochastic processes and increased computational cost. While acceptable in offline analysis, further investigation is required to assess scalability and efficiency in real-time or resource-constrained clinical environments. Overall, while these limitations highlight opportunities for future improvement, they do not detract from the central contribution of this work. The proposed NiOA–optimized DeepGBM framework represents a robust and clinically relevant advancement in OCD-related predictive modeling and provides a strong foundation for future research in intelligent mental health assessment systems.

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

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

6 Conclusion and Future Work

This study presented a comprehensive investigation into the application of advanced deep learning and metaheuristic optimization techniques for OCD-related classification using structured demographic and clinical data. A systematic experimental framework was designed to evaluate multiple deep learning models and to examine the impact of metaheuristic-driven hyperparameter optimization on predictive performance. The results clearly demonstrate that integrating the Ninja Optimization Algorithm with DeepGBM leads to substantial and consistent performance gains across all evaluated metrics.

The NiOA-optimized DeepGBM model exhibited marked improvements in classification accuracy, sensitivity, specificity, predictive values, and overall F-score when compared with both baseline models and alternative optimized variants. These gains highlight the effectiveness of NiOA in navigating complex, high-dimensional hyperparameter spaces and in identifying configurations that enhance both predictive accuracy and generalization capability. The robustness of the optimized model, reflected in its balanced performance across all metrics, confirms that the proposed optimization framework successfully addresses key limitations observed in non-optimized deep learning models, such as sensitivity to hyperparameter selection and susceptibility to suboptimal convergence.

The findings of this study have important practical implications for clinical practice and mental health research. The high sensitivity and specificity achieved by the NiOA-optimized DeepGBM model indicate its potential utility as a reliable decision-support tool for clinicians and mental health professionals. By providing accurate and consistent classification outcomes, the proposed framework can assist in reducing diagnostic uncertainty, supporting early identification of OCD-related patterns, and complementing traditional clinical assessments.

Moreover, the robustness and scalability of the optimized model make it a strong candidate for integration into intelligent mental health assessment systems. Such systems could leverage automated predictive analytics to support personalized care pathways, optimize resource allocation, and enhance longitudinal monitoring of patients. The use of structured clinical and demographic data further facilitates adoption in real-world healthcare environments, where such information is routinely collected and maintained.

Despite the promising outcomes achieved in this study, several avenues for future research remain open. One important direction involves the development of advanced hybrid metaheuristic designs that combine the strengths of NiOA with complementary optimization strategies. Such hybrids may further enhance convergence speed, stability, and scalability, particularly when dealing with increasingly large and complex datasets.

Another potential extension of this work lies in expanding the classification framework to support multi-class OCD severity prediction. Moving beyond binary classification toward fine-grained severity levels would provide more clinically actionable insights and enable more personalized treatment planning. This extension would require careful adaptation of both the modeling and optimization components to accommodate multi-class objectives.

Finally, future studies should investigate the feasibility of real-time deployment of the proposed framework within clinical decision support platforms. Addressing challenges related to computational efficiency, data privacy, and system integration will be essential for translating the research outcomes into practical, real-world applications. Exploring these directions will further strengthen the contribution of optimized deep learning models to intelligent mental healthcare systems and broaden their impact in clinical practice.

Data Availability

The dataset used in this study is publicly available on Kaggle at https://www.kaggle.com/datasets/ohinhaque/ocd-patient-dataset-demographics-and-clinical-data/data.

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

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

Declarations

- **Acknowledgments**
  Not applicable.
- **Conflict of interest/Competing interests**
  The authors declare that they have no conflicts of interest to report regarding the present study.
- **Ethics approval and consent to participate**
  Not applicable.
- **Consent for publication**
  Not applicable.
- **Funding**
  No Fund

References

[1] W. H. Organization, *World mental health report: Transforming mental health for all*. World Health Organization, 2022.

[2] W. H. Organization, *Optimizing brain health across the life course: WHO position paper*. World Health Organization, 2022.

[3] A. Dondu and L. Sevincok, "Clinical characteristics of obsessive-compulsive disorder comorbid with obsessive-compulsive personality disorder: Subtype implications," *Frontiers in psychiatry*, vol. 16, p. 1 577 042, 2025. DOI: https://doi.org/10.3389/fpsyt.2025.1577042.

[4] S. Pardossi, A. Cuomo, and A. Fagiolini, "Unraveling the boundaries, overlaps, and connections between schizophrenia and obsessive–compulsive disorder (ocd)," *Journal of Clinical Medicine*, vol. 13, no. 16, 2024, ISSN: 2077-0383. DOI: 10.3390/jcm13164739. [Online]. Available: https://www.mdpi.com/2077-0383/13/16/4739.

[5] A. Borrego-Ruiz and J. J. Borrego, "Biological, psychosocial, and microbial determinants of childhood-onset obsessive–compulsive disorder: A narrative review," *Children*, vol. 12, no. 8, 2025, ISSN: 2227-9067. DOI: 10.3390/children12081063. [Online]. Available: https://www.mdpi.com/2227-9067/12/8/1063.

[6] J. Kim et al., "Artificial intelligence in obsessive-compulsive disorder: A systematic review," *Current Treatment Options in Psychiatry*, vol. 12, no. 1, p. 23, 2025. DOI: https://doi.org/10.1007/s40501-025-00359-8.

[7] N. Kansara, B. Vala, and M. S. Shaikh, "Machine learning in mental health: Bridging gaps in diagnosis and intervention," in *2025 Eleventh International Conference on Bio Signals, Images, and Instrumentation (ICBSII)*, 2025, pp. 1–9. DOI: 10.1109/ICBSII65145.2025.11013436.

[8] B. Liu, C. Hu, and P. Bao, "Precision tms through the integration of neuroimaging and machine learning: Optimizing stimulation targets for personalized treatment," *Frontiers in Human Neuroscience*, vol. 19, p. 1 682 852, 2025. DOI: https://doi.org/10.3389/fnhum.2025.1682852.

[9] S. Kulkarni, "Expert systems in behavioral and mental healthcare: Applications of ai in decision-making and consultancy," *Big Data Analytics and Machine Intelligence in Biomedical and Health Informatics: Concepts, Methodologies, Tools and Applications*, pp. 147–186, 2022. DOI: https://doi.org/10.1002/9781119792376.ch6.

[10] B. Stahnke, "A systematic review of misdiagnosis in those with obsessive-compulsive disorder," *Journal of Affective Disorders Reports*, vol. 6, p. 100 231, 2021, ISSN: 2666-9153. DOI: https://doi.org/10.1016/j.jadr.2021.100231. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666915321001578.

[11] L. Weinberg, L. A. Martin, K. M. Post, and E. J. Ricketts, "Psychologists' diagnostic accuracy and treatment recommendations for obsessive-compulsive disorder," *Journal of Clinical Psychology*, vol. 81, no. 5, pp. 324–333, 2025. DOI: https://doi.org/10.1002/jclp.23775.

[12] X. Liu and Q. Fan, "Early identification and intervention in pediatric obsessive-compulsive disorder," *Brain Sciences*, vol. 13, no. 3, 2023, ISSN: 2076-3425. DOI: 10.3390/brainsci13030399. [Online]. Available: https://www.mdpi.com/2076-3425/13/3/399.

[13] D. Faustino, M. M. Gonçalves, R. Braga, M. J. Faria, and J. T. Oliveira, "Exposure and response prevention in ocd: A framework to capitalize change," *Journal of Clinical Psychology*, 2025. DOI: https://doi.org/10.1002/jclp.23797.

[14] S. C. Jessup, A. Mariaskin, and B. O. Olatunji, "Strategies for optimizing traditional exposure and response prevention: A case study example in an adolescent with contamination-based ocd," *Journal of Clinical Psychology*, vol. 81, no. 3, pp. 182–192, 2025. DOI: https://doi.org/10.1002/jclp.23758.

[15] 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. DOI: https://doi.org/10.3389/fpsyt.2025.1750938.

[16] 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. DOI: https://doi.org/10.1186/s12888-023-05299-2.

[17] 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. 100 089, 2024, ISSN: 2950-0044. DOI: https://doi.org/10.1016/j.xjmad.2024.100089. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2950004424000439.

[18] M. Sachdeva, H. K. Sharma, A. Kumar, A. Bansal, K. Saluja, and Shikha, "Applying neural imaging and ml to ocd severity prediction," in *2022 Seventh International Conference on Parallel, Distributed and Grid Computing (PDGC)*, 2022, pp. 402–406. DOI: 10.1109/PDGC56933.2022.10053257.

[19] 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, 2025, ISSN: 2077-0383. DOI: 10.3390/jcm14072442. [Online]. Available: https://www.mdpi.com/2077-0383/14/7/2442.

[20] W. S. Aljebreen, N. A. Alturaiqi, M. I. Almushyti, H. F. Alhasson, and S. S. Alharbi, "Advancing precision psychiatry: Machine learning integration with neuroimaging for early detection and diagnosis of obsessive-compulsive disorder," *Psychiatry Research: Neuroimaging*, vol. 357, p. 112 137, 2026, ISSN: 0925-4927. DOI: https://doi.org/10.1016/j.pscychresns.2026.112137. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925492726000028.

[21] M. Naderi and A. Jahanian-Najafabadi, "A systematic review of eeg-based machine learning classifications for obsessive-compulsive disorder: Current status and future directions," *BMC psychiatry*, vol. 25, no. 1, p. 854, 2025. DOI: https://doi.org/10.1186/s12888-025-07296-z.

[22] U. M. Haque, E. Kabir, and R. Khanam, "Early detection of paediatric and adolescent obsessive–compulsive, separation anxiety and attention deficit hyperactivity disorder using machine learning algorithms," *Health information science and systems*, vol. 11, no. 1, p. 31, 2023. DOI: https://doi.org/10.1007/s13755-023-00232-z.

[23] K. Patel, A. K. Tripathy, L. N. Padhy, S. K. Kar, S. K. Padhy, and S. P. Mohanty, "Accu-help: A machine-learning-based smart healthcare framework for accurate detection of obsessive compulsive disorder," *SN Computer Science*, vol. 5, no. 1, p. 36, 2023. DOI: https://doi.org/10.1007/s42979-023-02380-1.

[24] Naseerullah, M. Hayat, N. Iqbal, M. Tahir, S. A. AlQahtani, and A. M. Alamri, "Early diagnosis of obsessives-compulsive disorder through gene expression analysis using machine learning models," *Chemometrics and Intelligent Laboratory Systems*, vol. 248, p. 105 107, 2024, ISSN: 0169-7439. DOI: https://doi.org/10.1016/j.chemolab.2024.105107. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169743924000479.

[25] M. Tubío-Fungueiriño et al., "Prediction of pharmacological response in ocd using machine learning techniques and clinical and neuropsychological variables," *Spanish Journal of Psychiatry and Mental Health*, vol. 18, no. 1, pp. 51–57, 2025, ISSN: 2950-2853. DOI: https://doi.org/10.1016/j.sjpmh.2024.11.001. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S295028532400070X.

[26] 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, ISSN: 0165-0327. DOI: https://doi.org/10.1016/j.jad.2024.01.157. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165032724001757.

[27] 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, ISSN: 0165-0327. DOI: https://doi.org/10.1016/j.jad.2021.09.042. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165032721010016.

[28] Z. Zhao et al., "Exploring potential resting-state eeg biomarkers of obsessive-compulsive disorder based on explainable machine learning analysis of independent training and test samples," *BMC psychiatry*, vol. 25, no. 1, p. 1164, 2025. DOI: https://doi.org/10.1186/s12888-025-07583-9.

[29] 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. DOI: 10.1109/TAFFC.2024.3395117.

[30] 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. DOI: https://doi.org/10.1038/s41380-023-02077-0.

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

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