EMITTER International Journal of Engineering Technology
Vol. 14, No. 1, June 2026, pp. 159~183
DOI : 10.24003/emitter.v14i1.1056

Particle Swarm Optimization–Tuned Random Forest for Multi-Class Mental Health Sentiment Classification from Textual Data

Supriady[1], Ode Andi Alamsyah[2], Nesya Salma Ramadhani[2], Syafrial Fachri Pane[2]

[1]School of Information Technology, Associate Degree Program of Informatics Engineering, Universitas Logistik dan Bisnis Internasional
[2]School of Information Technology, Applied Bachelor Program of Informatics Engineering, Universitas Logistik dan Bisnis Internasional
Corresponding Author: supriady@ulbi.ac.id

Received January 10, 2026; Revised March 13, 2026; Accepted June 10, 2026

Abstract

Mental health sentiment classification from textual data has attracted increasing attention as a computational approach to support large-scale psychological assessment; however, multi-class classification remains challenging due to noisy text, class imbalance, and semantic overlap among categories. This study proposes and evaluates a machine learning framework for seven-class mental health sentiment classification that integrates enhanced text preprocessing with lemmatization, data augmentation via back-translation, TF-IDF feature extraction, and systematic model evaluation across multiple classifiers, including Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, K-Nearest Neighbors, AdaBoost, and XGBoost, under three hyperparameter tuning strategies: Grid Search, Random Search, and Particle Swarm Optimization (PSO). Experimental results indicate that ensemble-based models consistently outperform single classifiers, with the PSO-optimized Random Forest achieving the best numerical performance, attaining an accuracy of 0.933, a macro F1-score of 0.923, and a ROC AUC of 0.989, demonstrating strong generalization and balanced class-level performance despite dataset imbalance. These findings confirm that the combination of robust preprocessing and metaheuristic-based hyperparameter optimization significantly enhances multi-class mental health sentiment classification and supports its potential use as a scalable decision-support tool for large-scale mental health screening, while not intended to replace clinical diagnosis.

Keywords: Computational mental health, Multi-class sentiment analysis, Ensemble classifiers, Hyperparameter optimization, Text-based machine learning

1. INTRODUCTION

The rising prevalence of mental health disorders, including depression, anxiety, bipolar disorder, stress, suicidal ideation, and personality disorders, constitutes a critical public health crisis. These conditions significantly impair individual functioning and social productivity, yet a substantial gap persists in global access to adequate psychiatric care [1], [2]. Factors including social stigma, high treatment costs, and the subjective nature of psychological assessment often delay early detection, reinforcing the need for scalable and objective support mechanisms[3]. Consequently, scalable and automated approaches for early mental health detection are increasingly required.

The widespread adoption of digital communication platforms has enabled individuals to express psychological states in writing. Content generated on social media, online forums, and digital interaction platforms contains linguistic patterns that reflect emotional conditions and mental distress[4], [5]. These characteristics position textual data as a valuable resource for automated mental health analysis. In this context, sentiment analysis supported by Natural Language Processing provides a computational approach for extracting mental health indicators from unstructured text [6], [7], [8].

Various machine learning approaches have been applied to mental health text classification in prior studies [9], [10], [11]. Despite reported effectiveness, most existing work focuses on binary or limited-class classification settings, which do not adequately represent the complexity and semantic overlap among different mental health conditions. Linguistic similarities across categories reduce discriminative capability, making multi-class mental health sentiment classification a challenging and comparatively underexplored research problem [12].

Model performance in text-based mental health classification is strongly influenced by hyperparameter configuration, particularly for ensemble-based learning approaches. Conventional hyperparameter tuning methods are commonly adopted but often exhibit inefficiency when applied to high-dimensional and complex parameter spaces [13]. Such limitations may lead to suboptimal parameter configurations, especially for ensemble models that are sensitive to structural and feature-related parameters [14].

Particle Swarm Optimization is a population-based metaheuristic algorithm inspired by collective swarm behavior and has demonstrated strong capability in exploring large and non-linear search spaces efficiently [15], [16]. Despite its effectiveness across various optimization problems, its application to hyperparameter tuning for ensemble learning in multi-class mental health sentiment analysis remains limited. Furthermore, only a small number of studies provide systematic comparisons between metaheuristic-based optimization and conventional tuning strategies within a unified experimental framework [17].

Text-based mental health datasets present additional challenges, including noisy textual content, sparse feature representations, and skewed

class distributions [18]. These characteristics can lead to biased predictions and misleading performance evaluations if not carefully addressed. Although various class balancing techniques exist, this study does not introduce an explicit data balancing method. Instead, it emphasizes robust text preprocessing, systematic hyperparameter optimization, and comprehensive evaluation using class-sensitive performance metrics to ensure reliable assessment across all mental health categories [19].

To address the identified research gaps, this study proposes a Random Forest model tuned using Particle Swarm Optimization for multi-class mental health sentiment classification based on textual data. The proposed framework integrates comprehensive text preprocessing, tokenization, stopword removal, data augmentation, feature vectorization, and rigorous model training. In addition, a set of machine learning classifiers comprising Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, K-Nearest Neighbors, AdaBoost, and XGBoost is evaluated using three hyperparameter tuning strategies, namely Grid Search, Random Search, and Particle Swarm Optimization, within a unified experimental framework to enable a fair and systematic comparison of model performance and optimization effectiveness [13], [14], [15], [16], [17].

This study has four contributions. First, it introduces a comprehensive framework for seven-class mental health sentiment classification that reflects the diversity of psychological conditions encountered in real-world textual data. Second, it demonstrates that metaheuristic-based optimization can effectively enhance classification performance compared to conventional tuning strategies [15], [16]. Third, it provides a thorough experimental evaluation using multiple performance metrics, offering deeper insights into per-class behavior, model robustness, and generalization in practical mental health text analysis scenarios [19]. Fourth, it presents a deployment-oriented integration of the trained model into a web-based application, enabling real-time inference, user interaction, and practical decision-support functionality beyond offline experimentation.

2. RELATED WORKS

Text-based analysis has become a prominent approach in mental health research, driven by the increasing availability of large-scale textual data from social media and online communication platforms [4], [5], [6], [7]. Previous studies indicate that linguistic features embedded in text can reflect psychological signals related to various mental health conditions. As a result, automated mental health classification using textual data has been explored as a computational strategy to support large-scale screening and early detection [3], [13].

Early research in this area predominantly employed conventional machine learning classifiers combined with bag-of-words or TF-IDF feature representations [7], [11], [12]. While these approaches yield reasonable performance in binary or restricted multi-class settings, their effectiveness

decreases when applied to more complex classification scenarios. Class imbalance and overlapping linguistic characteristics among mental health categories limit discriminative capability, motivating the exploration of more robust modeling strategies [18].

To overcome these limitations, ensemble-based learning methods have gained increasing attention. Approaches based on Random Forest, Gradient Boosting, bagging-based ensembles, and hybrid ensemble architectures have been shown to enhance robustness and generalization through the aggregation of multiple weak learners [5], [9], [10]. Previous research highlights that Random Forest-based approaches are particularly effective in mental health prediction tasks, owing to their effectiveness in managing high-dimensional feature spaces, noisy data, and non-linear relationships [9], [11]. Ensemble and hybrid models have also been reported to outperform individual classifiers in various mental health prediction scenarios [5], [10].

Hyperparameter optimization has been identified as a crucial component in improving model performance. Conventional tuning strategies, including Grid Search and Random Search, are commonly adopted due to their simplicity and reliability [13]. However, these approaches often require substantial computational resources and may fail to efficiently explore large parameter spaces, especially for ensemble-based models [14]. Consequently, stochastic metaheuristic algorithms, such as Particle Swarm Optimization and other evolutionary algorithms, are increasingly explored for machine learning optimization tasks [15], [16].

Recent advances in machine learning have established Particle Swarm Optimization (PSO) as one of the most widely adopted metaheuristic algorithms for hyperparameter optimization because of its ability to efficiently explore complex and high-dimensional search spaces while maintaining relatively low computational cost [15], [16]. Recent studies have successfully applied PSO to optimize Support Vector Machines, Random Forest, Gradient Boosting, and Neural Network models, frequently achieving superior predictive performance compared with conventional optimization approaches such as Grid Search and Random Search [20], [21], [22]. These findings demonstrate that PSO has become an effective optimization technique across a broad range of machine learning applications.

In the field of mental health sentiment classification, however, the application of PSO remains relatively limited. Existing studies primarily focus on binary classification tasks such as depression detection or suicide risk prediction, or they optimize a single classifier without systematically comparing different optimization strategies under identical experimental settings [17], [23]. Meanwhile, recent mental health classification research has increasingly emphasized deep learning architectures and large language models [19], [24], leaving comparatively few studies that investigate lightweight ensemble-based machine learning models enhanced through swarm-intelligence optimization. Consequently, the current state of the art indicates that comprehensive evaluations integrating PSO, multiple machine

learning classifiers, conventional hyperparameter tuning methods, and multi-class mental health sentiment classification remain scarce.

Several studies suggest that Particle Swarm Optimization can effectively guide the search process toward optimal or near-optimal hyperparameter configurations, leading to improved classification performance [15], [16], [17]. Nevertheless, most existing research applying PSO in mental health-related machine learning focuses on binary classification problems, specific populations, or a limited number of mental health categories [9], [12]. Comprehensive evaluations that combine ensemble-based classifiers, multiple tuning strategies, and multi-class mental health text classification remain relatively scarce [19], [25].

Based on these observations, there is a clear need for a systematic investigation that evaluates the impact of Particle Swarm Optimization on ensemble-based models, particularly Random Forest, within a multi-class mental health sentiment classification framework. This study addresses this gap by providing a comprehensive comparison across multiple classifiers, tuning strategies, and evaluation metrics using large-scale textual mental health data.

3. ORIGINALITY

Previous studies on text-based mental health status classification exhibit wide variation in datasets, modeling strategies, and analytical depth, yet several recurring limitations remain. Binary or limited-class frameworks are still prevalent; for example, Ghosal and Jain employ fastText and TF-IDF features with an XGBoost classifier to distinguish depression and suicide content on social media, which limits applicability to real-world scenarios involving multiple coexisting mental health conditions [23]. To expand label coverage, Sutranggono et al. propose a hierarchical multi-class framework, while Hossain et al. introduce hybrid deep learning architectures integrating opinion modeling and multi-task learning; however, these approaches increase computational complexity and reduce scalability and reproducibility [24], [26]. In parallel, Kallstenius et al. demonstrate that feature-engineered traditional NLP models remain competitive against large language models, yet their analysis primarily emphasizes benchmarking rather than systematic ensemble optimization [19]. Juanita et al. further explore hybrid machine learning optimized via Genetic Algorithms and Bagging for three-class clinical classification, but do not investigate swarm-based optimization or extended label diversity [27].

Collectively, existing studies emphasize restricted label spaces, architectural complexity, or benchmarking-oriented evaluation, leaving the systematic optimization of lightweight ensemble models combined with enhanced linguistic preprocessing and data augmentation for scalable multi-class classification insufficiently explored. Addressing this gap, the present study introduces a computationally efficient seven-class mental health status classification framework that integrates enhanced preprocessing (including

lemmatization and back-translation augmentation), TF-IDF feature representation, and Particle Swarm Optimization–driven ensemble learning, and a web-based deployment architecture that supports real-time inference and practical decision-support deployment.

4. SYSTEM DESIGN

Figure 1 shows the architecture of the proposed system. The framework follows a sequential pipeline that transforms raw textual statements into multi-class mental health predictions through preprocessing, data augmentation, feature extraction, machine learning model training, hyperparameter optimization, and comprehensive evaluation. This architecture ensures modularity and enables a fair comparison across different models and optimization strategies.

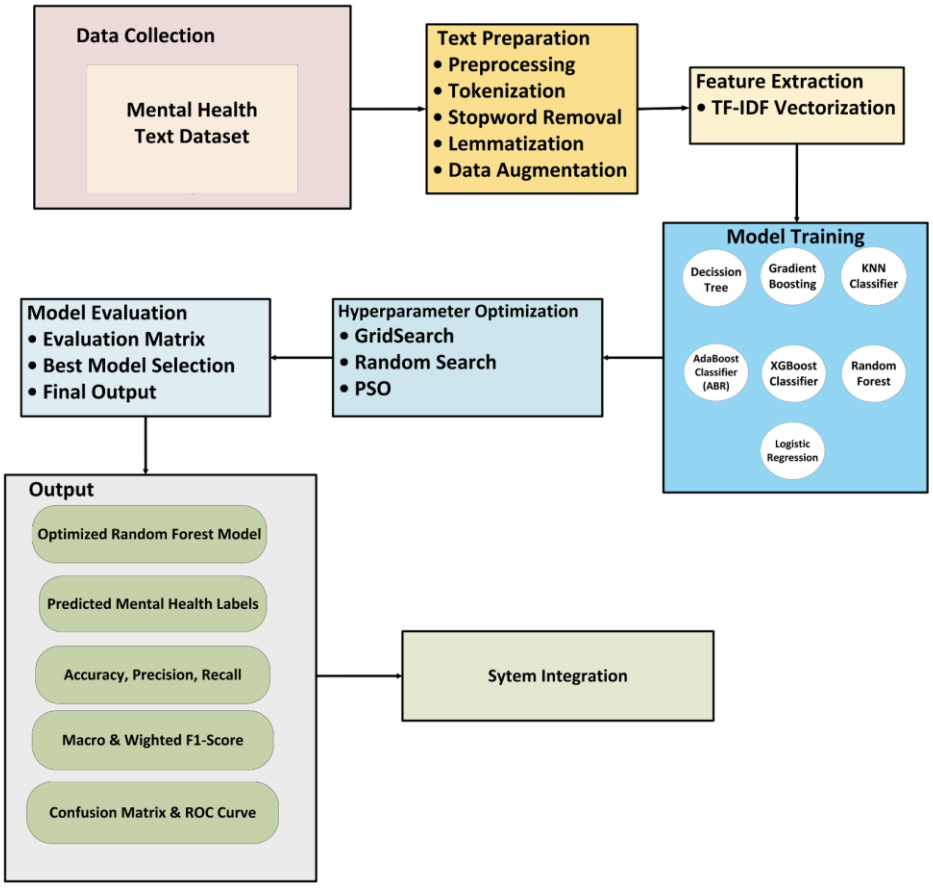

Figure 1. Architecture of the proposed system

4.1 Data Collection

The system utilizes a publicly available mental health sentiment dataset obtained from Kaggle and entitled Sentiment Analysis for Mental Health, released in 2024. The dataset consists of curated textual statements representing emotional expressions, thoughts, and psychological conditions aggregated from multiple public sources, including social media platforms,

Reddit discussions, Twitter content, and chatbot conversations. Each text instance is assigned a single mental health label under a single-label classification scheme. Such unstructured textual data inherently contain noise, non-standard language patterns, and sparse feature representations, which pose challenges for feature extraction and downstream classification in mental health analysis [28]. Prior to further processing, the dataset was validated to ensure completeness and label consistency [3], [7].

4.2 Text Preparation

Text preparation combines text preprocessing and data augmentation. Textual preprocessing includes lowercase normalization and the removal of punctuation, special characters, and numerical tokens. This is followed by tokenisation, stopword removal and lemmatisation to standardise the textual input.

The application of data augmentation to the training set is used to enrich textual variation and improve model generalisation. The augmentation technique employed was back-translation, where each textual instance was first translated from English to French and then subsequently translated back into English using the TextBlob library. This approach introduces lexical and syntactic variations while preserving the original semantic meaning of the text. After augmentation, the same preprocessing pipeline was reapplied to ensure consistency across all samples. To prevent data leakage, no augmented samples were included in the test set [29], [30].

4.3 Feature Extraction

Textual data are transformed into numerical feature vectors using the Term Frequency–Inverse Document Frequency (TF-IDF) weighting scheme [31]. TF-IDF measures the relative importance of a term within a document by balancing its local frequency against its global distribution across the entire corpus, thereby emphasizing discriminative terms while reducing the influence of commonly occurring words. The TF-IDF weight is computed as follows:

$$\mathrm{TF-IDF(t,d)} = \mathrm{TF(t,d)} \times \log\left(\frac{\mathrm{N}}{\mathrm{DF(t)}}\right) \quad (1)$$

In Equation (1), t denotes a term (word or token), and d represents a document in the corpus. $TF(t,d)$ is the term frequency, defined as the number of occurrences of term t in document d. $DF(t)$ refers to the document frequency, indicating the number of documents in the corpus that contain term t. N represents the total number of documents in the corpus. The logarithmic factor $\log\left(\frac{\mathrm{N}}{\mathrm{DF(t)}}\right)$ corresponds to the inverse document frequency, which assigns higher weights to terms that appear in fewer documents and lower weights to terms that are widely distributed across the corpus. Consequently, the TF-IDF value reflects the importance of a term in a specific document

relative to the overall corpus distribution and serves as an effective feature representation for text classification tasks [7].

4.4 Model Training and Optimization

Multiple machine learning classifiers are employed in this study, comprising Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, K-Nearest Neighbors, AdaBoost, and XGBoost. These classifiers represent a diverse range of linear, non-linear, and ensemble learning architectures commonly used in text classification and mental health predictive analytics [2], [17].

Model training and hyperparameter optimization are conducted jointly using three tuning strategies, namely Grid Search, Random Search, and Particle Swarm Optimization[20], [21], [32]. The first two strategies serve as conventional baseline tuning methods, while Particle Swarm Optimization is employed as a metaheuristic approach to efficiently explore the hyperparameter search space.

Particle Swarm Optimization (PSO) was implemented using a swarm of 10 particles and a maximum of 20 iterations. An inertia weight of 0.9 was adopted, whereas the cognitive and social acceleration coefficients were set to c1 = 0.6 and c2 = 0.3, respectively. These coefficients were selected through preliminary experiments that compared three acceleration coefficient configurations, namely (2.0, 2.0), (1.0, 1.0), and (0.6, 0.3), before the final optimization process. The selected configuration was then applied consistently throughout all PSO-based hyperparameter tuning experiments. This parameter selection strategy is consistent with previous studies reporting that PSO performance is highly dependent on its control parameter configuration and that suitable parameter values should be determined empirically according to the characteristics of the optimization problem [33], [34].

The optimization objective was formulated as minimizing a loss function based on the average classification accuracy obtained through five-fold cross-validation on the training data (loss = 1 − accuracy). This cross-validation-based optimization strategy was adopted to ensure robust and stable hyperparameter selection while reducing the risk of overfitting, particularly for ensemble-based models with high-dimensional and complex parameter spaces [13], [15].

4.5 Model Evaluation

Predictive performance is assessed by using multiple evaluation metrics to provide a comprehensive analysis. Accuracy is reported as an overall performance indicator, while precision, recall, and F1-score are employed to analyse performance at class level. Macro and weighted F1-scores, confusion matrix analysis, and ROC curve evaluation are employed to assess classification effectiveness across all mental health categories [19], [24].

The evaluation metrics are defined as follows:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (3)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (4)$$

$$\text{F1-score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

In Equation (2)–(5), TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively. For multi-class evaluation, macro and weighted F1-scores are computed by aggregating class-wise F1-scores with equal and class frequency-based weighting, respectively. Confusion matrix analysis and ROC curve evaluation are further employed to examine misclassification patterns and discrimination capability across all mental health categories [19].

4.6 Pseudocode

Algorithm 1 presents the pseudocode of the proposed Particle Swarm Optimization–optimized Random Forest framework for multi-class mental health sentiment classification. The algorithm summarizes the main stages of data preparation, feature extraction, hyperparameter optimization, and final model training.

Algorithm 1: End-to-End Pipeline of PSO-Optimized Random Forest for Multi-Class Mental Health Sentiment Classification

Input: Raw text dataset $\text{D} = (x_i, y_i), y_i$ {Normal, Depression, Anxiety, Stress, Bipolar, Suicidal, Personality Disorder}

Output: Optimal hyperparameters θ^*, best accuracy Acc^*, predicted labels $\hat{y}_{test}$

Phase 1: Data Preparation

Step 1: Text Preprocessing

For each $x_i \in D$ **do**

1. Convert text to lowercase;
2. Remove punctuation, numbers, and special characters;
3. Perform tokenization;
4. Remove stopwords;
5. Apply lemmatization;

Step 2: Text Augmentation and Normalization

6. Generate augmented texts D_{aug} using text augmentation techniques;
7. Reapply text preprocessing on D_{aug};
8. Merge datasets $D_{final} \leftarrow D \cup D_{aug}$;

Step 3: Feature Extraction

9. Transform preprocessed texts in D_final into feature vectors X using TF-IDF vectorization;
10. Step 4: Dataset Splitting
11. Split (X, y) into training set (X_{train}, y_{train}) and testing set (X_{test}, y_{test});

Phase 2: PSO-Based Hyperparameter Optimization

Step 5: Swarm Initialization

12. Initialize swarm P with N particles;
13. Each particle represents a hyperparameter set θ_i;
14. Initialize particle positions x_i and velocities v_i;
15. Set $pbest_i \leftarrow x_i$, gbest $\leftarrow$ arg max(Fitness);

Step 6: PSO Optimization Loop

While iter < Max_Iterations **do**

For each $particle\ i \in P$ **do**

17. Decode x_i into θ_i;
18. Train Random Forest model M_i on (X_{train}, y_{train});
19. Predict validation labels $\hat{y}_{val}$;
20. Compute fitness using Accuracy or F1-Macro;
21. **if** $Fitness_i >$ Fitness($pbest_i$) **then**
22. $pbest_i \leftarrow x_i$;
23. **if** $Fitness_i >$ Fitness ($gbest_i$) **then**
24. $gbest_i \leftarrow x_i$;

For each $particle\ i \in P$ **do**

25. vi $\leftarrow wv_i + c_1 r_1(pbest_i - x_i) + c_2 r_2(\text{gbest} - x_i)$;
26. $x_i \leftarrow x_i + v_i$;

Phase 3: Final Modeling

27. $\theta^* \leftarrow$ gbest;
28. Train final Random Forest model M^*on (X_{train}, y_{train});
29. Predict $\hat{y}_{test} \leftarrow M^*.predict(X_{test})$
30. Compute final evaluation metrics (Accuracy, Precision, Recall, F1-score);
31. Return θ^*, Acc^*, $\hat{y}_{test}$

4.7 Web-Based System Integration

The trained PSO-optimized Random Forest model was integrated into a web-based application to enable real-time mental health status classification and interactive user access. This deployment demonstrates the practical feasibility of extending the proposed framework beyond offline experimentation toward a lightweight decision-support system.

The system adopts a client–server architecture in which textual input is submitted through the web interface and transmitted to the backend service for processing. Automated preprocessing, including normalization and lemmatization, is performed prior to feature transformation using the trained TF-IDF vectorizer. The resulting feature vectors are then passed to the PSO-optimized Random Forest model for inference, and the predicted class label

along with its probability distribution is returned to the frontend for visualization.

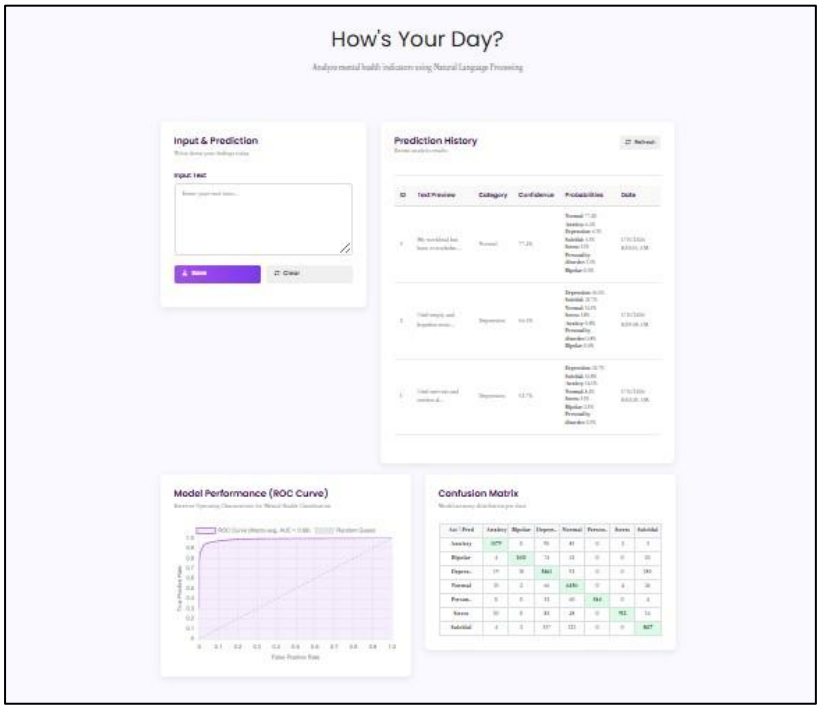

Figure 2. Web-based interface of the integrated mental health classification system.

As illustrated in Figure 2, the user interface provides an input panel for text submission, a prediction history table displaying previously analyzed texts with predicted categories and confidence scores, and visualization components for model performance, including ROC curves and confusion matrices. These components support transparency, traceability, and usability in the prediction process while facilitating exploratory analysis and result interpretation.

5. EXPERIMENT AND ANALYSIS

This section describes the experimental configuration, dataset characteristics, evaluation metrics, and performance analysis of the proposed multi-class mental health sentiment classification framework. The experiments are designed to assess the performance of multiple machine learning models under different hyperparameter optimization strategies, with particular emphasis on the effect of Particle Swarm Optimization on classification performance.

5.1 Data Description

The experiments conducted in this work used a mental health sentiment dataset from Kaggle titled Sentiment Analysis for Mental Health. The dataset is constructed through the aggregation and curation of textual data from

multiple publicly available sources, resulting in a comprehensive collection of mental health–related statements [35]. The data compilation process involves cleaning, filtering, and unifying text samples to support robust sentiment analysis and multi-class classification tasks.

The dataset integrates textual data derived from several Kaggle sources, including chatbot conversations, online mental health discussions, and social media content related to psychological well-being[36]. These sources represent a wide range of psychological contexts. The integration of multiple data sources enhances linguistic diversity and enables the dataset to capture varied patterns of mental health expression [37].

Each text instance is annotated with one of seven predefined mental health categories. The dataset follows a single-label annotation scheme, in which each statement is associated with exactly one mental health label. Accordingly, the classification task is formulated as a single-label multi-class problem [38].

The original dataset contains 53,043 text samples, which increase to 106,860 samples after text preprocessing and data augmentation. Data augmentation is applied exclusively to the training set using back-translation, while the test set remains unchanged to preserve unbiased evaluation. All textual data are written in English. The dataset is divided into training and testing partitions using an 80:20 split ratio, where the training data are used for model training and hyperparameter optimization, and the testing data are reserved for performance evaluation.

The dataset consists of three primary attributes. Each entry has unique identifier, along with a textual statement and a corresponding mental health label. To adhere to ethical protocols, the dataset is anonymized and does not contain personally identifiable information, ensuring its viability for computational mental health analysis [36].

Table 1 summarizes the key characteristics of the mental health sentiment dataset used in this study, including data sources, class categories, sample size, and feature representation.

5.2 Class Distribution Analysis

Figure 3 shows the distribution of samples across the seven mental health categories in the test dataset. The distribution is clearly imbalanced, with the normal and depression classes containing the largest number of samples, while personality disorder and bipolar disorder represent minority classes. This imbalance reflects real-world mental health data characteristics and increases the difficulty of multi-class classification, particularly for minority categories [39].

Table 1. key characteristics of the dataset

Attribute	Description
Dataset Source	Kaggle
Dataset Name	Sentiment Analysis for Mental Health
Dataset Type	Public aggregated text-based mental health dataset
Data Origin	Reddit posts, Twitter posts, chatbot conversations, and social media content
Language	English
Total Samples (Before Processing)	53,043
Total Samples (After Processing and Augmentation)	106,860
Number of Classes	7
Mental Health Categories	Normal, Depression, Suicidal, Anxiety, Stress, Bipolar Disorder, Personality Disorder
Labeling Schema	Single-label classification
Text Attributes	Unique ID, Statement, Mental Health Status
Train-Test Split	80:20
Feature Representation	TF-IDF

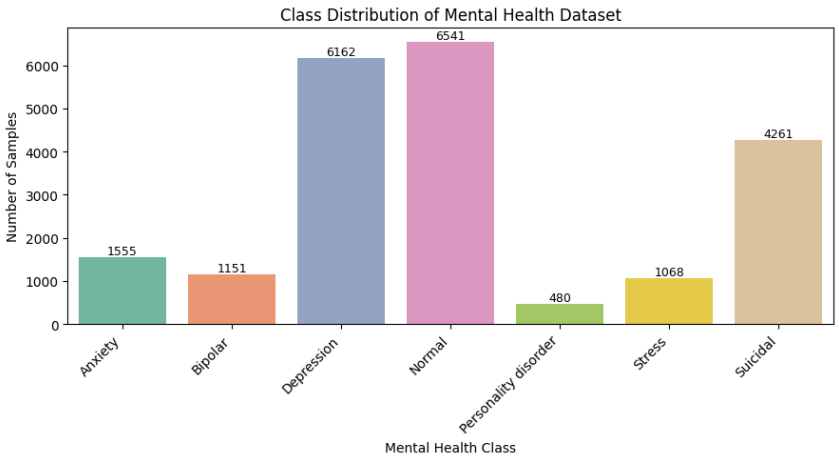

Figure 3. Class Distribution of Mental Health Dataset

Although the original dataset exhibited class imbalance, data augmentation was applied during preprocessing to enrich the minority classes and improve data diversity [40]. No additional algorithm-level balancing strategies, such as class weighting, were employed. Model performance was evaluated on the resulting data distribution using macro-averaged precision,

recall, F1-score, and ROC-AUC, as these metrics provide a more comprehensive assessment of multiclass imbalanced classification performance than accuracy alone [41].

5.3 Experimental Setup

In the experimental setup, textual inputs are first encoded into numerical representations using the Term Frequency Inverse Document Frequency approach. The experimental framework evaluates a set of machine learning classifiers. To ensure a fair and consistent comparison, each model is examined under four different configurations, including a default setting and three hyperparameter optimization strategies [42].

Hyperparameter optimization is conducted using the training data, while all reported performance results are obtained from the test data [22], [43]. Identical experimental conditions are applied across all models to ensure a fair comparison. The experiments are implemented using Python and standard machine learning libraries.

5.4 Evaluation Metrics

To provide a comprehensive performance evaluation, multiple metrics are employed. Accuracy is used as a general indicator of overall predictive performance. Precision, recall, and F1-score are calculated to analyze class-level behavior. Macro and weighted F1-scores are reported to account for class imbalance [39]. In addition, confusion matrix analysis and Receiver Operating Characteristic curve evaluation using a one-vs-rest strategy are conducted to examine prediction patterns and discrimination capability across all mental health categories.

5.5. Result and Discussion

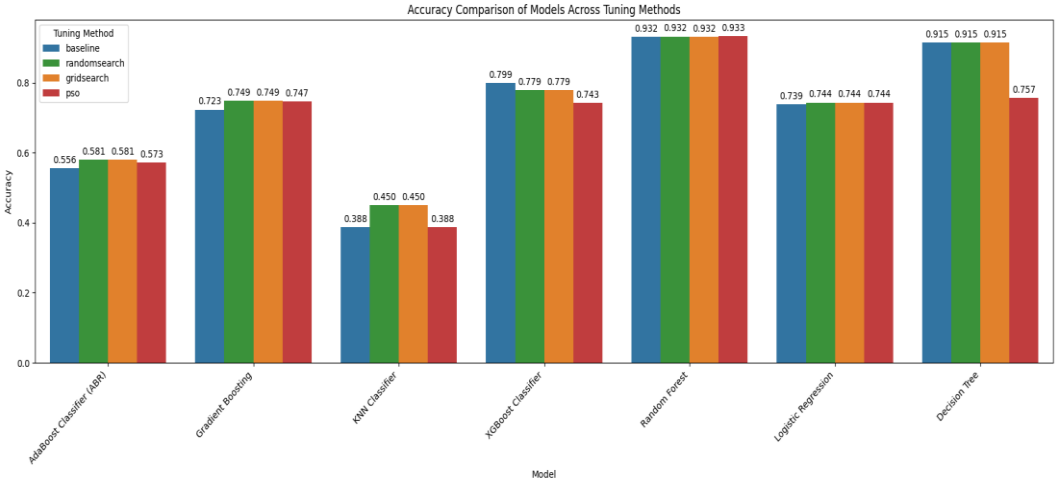

Figure 4. Accuracy Comparison of Models Across Tuning Methods

Figure 4 shows that ensemble-based models consistently outperform single classifiers across all tuning strategies. Among all evaluated approaches, the Random Forest optimized using Particle Swarm Optimization achieves the highest accuracy of 0.933. This result indicates that PSO effectively explores the hyperparameter space for complex ensemble models, while simpler models such as K-Nearest Neighbors gain limited benefit from metaheuristic tuning due to the high dimensionality of TF-IDF features [22], [44], [45].

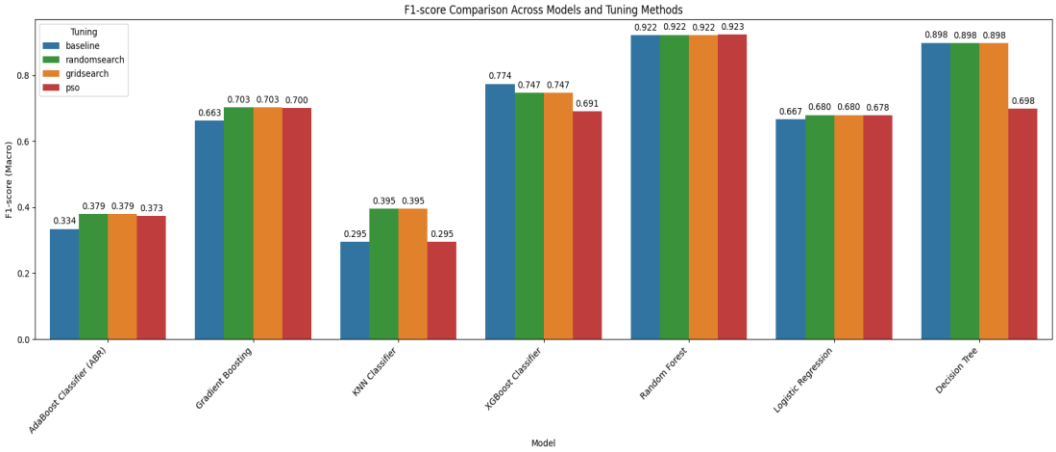

Figure 5. F1-Score Comparison Across Models and Tuning Methods

Figure 5 presents the macro F1-score comparison, highlighting the impact of hyperparameter tuning on class-level performance. The PSO-tuned Random Forest attains the highest macro F1-score of 0.923, demonstrating balanced performance across both majority and minority classes. In contrast, Decision Tree exhibits reduced F1-score under PSO tuning, suggesting that aggressive optimization may lead to overfitting in simpler models [43], [45].

Table 2 summarizes the performance of the evaluated models across different hyperparameter tuning strategies. The results are reported using multiple evaluation metrics, including accuracy, macro-averaged precision, recall, F1-score, and ROC AUC, computed under a one-vs-rest scheme to support evaluation of both overall effectiveness and class-level performance. As reported in this table, ensemble-based classifiers consistently outperformed single classifiers across all evaluation metrics. The PSO-optimized Random Forest achieved the highest overall performance, while the differences among the three hyperparameter tuning strategies for Random Forest were marginal, indicating that all three tuning strategies produced comparable results.

Table 2. performance comparison across models and tuning strategies.

Model	Tuning	Accuracy	Precision (Macro)	Recall (Macro)	F-1 Score (Macro)	ROC AUC (OvR)
AdaBoost	Baseline	0.556	0.650	0.326	0.334	0.789
Gradient Boosting	Baseline	0.723	0.765	0.613	0.663	0.916
KNN	Baseline	0.388	0.665	0.281	0.295	0.874
XGBoost	Baseline	0.799	0.842	0.703	0.774	0.955
Random Forest	Baseline	0.932	0.959	0.892	0.922	0.988
Logistic Regression	Baseline	0.739	0.737	0.628	0.669	0.924
Decision Tree	Baseline	0.915	0.913	0.886	0.898	0.948
AdaBoost	Random Search	0.581	0.702	0.359	0.379	0.820
Gradient Boosting	Random Search	0.749	0.75	0.656	0.703	0.930
KNN	Random Search	0.450	0.619	0.379	0.395	0.893
XGBoost	Random Search	0.779	0.831	0.698	0.747	0.947
Random Forest	Random Search	0.932	0.961	0.892	0.922	0.990
Logistic Regression	Random Search	0.744	0.723	0.650	0.680	0.926
Decision Tree	Random Search	0.915	0.913	0.885	0.898	0.948
AdaBoost	Grid Search	0.581	0.702	0.359	0.379	0.820
Gradient Boosting	Grid Search	0.749	0.787	0.656	0.703	0.930
KNN	Grid Search	0.450	0.619	0.379	0.395	0.893
XGBoost	Grid Search	0.779	0.831	0.698	0.747	0.947
Random Forest	Grid Search	0.932	0.960	0.891	0.922	0.990
Logistic Regression	Grid Search	0.744	0.723	0.650	0.680	0.926
Decision Tree	Grid Search	0.916	0.912	0.885	0.898	0.948
AdaBoost	PSO	0.573	0.692	0.352	0.373	0.816
Gradient Boosting	PSO	0.747	0.787	0.652	0.700	0.928
KNN	PSO	0.388	0.665	0.281	0.295	0.874
XGBoost	PSO	0.743	0.806	0.636	0.691	0.928
Random Forest	**PSO**	**0.933**	**0.960**	**0.893**	**0.923**	**0.989**
Logistic Regression	PSO	0.744	0.737	0.644	0.678	0.925
Decision Tree	PSO	0.759	0.815	0.639	0.698	0.888

To determine an appropriate PSO configuration, three commonly used acceleration coefficient settings were evaluated in preliminary experiments before conducting the main optimization process. As shown in Table 3, the configuration with c1 = 0.6 and c2 = 0.3 achieved the highest classification accuracy. Consequently, this configuration was adopted in all subsequent PSO-based hyperparameter tuning experiments.

Table 3. preliminary evaluation of PSO acceleration coefficient configurations.

c1	c2	Accuracy
2.0	2.0	0.932
1.0	1.0	0.932
0.6	**0.3**	**0.933**

Table 4 provides the statistical validation for our model comparisons using the McNemar test [46]. The performance of the PSO-based model did not differ significantly from either Grid Search (χ^2 = 3.314, p = 0.069) or Random Search (χ^2 = 0.405, p = 0.525). This implies the gains over other tuning strategies are not statistically significant at a 5% alpha level.

Table 4. Statistical comparison between hyperparameter tuning strategies using McNemar's test

Comparison	Statistic	P-value	Conclusion
PSO vs Grid Search	3.314	0.069	Not significant
PSO vs Random Search	0.405	0.525	Not significant

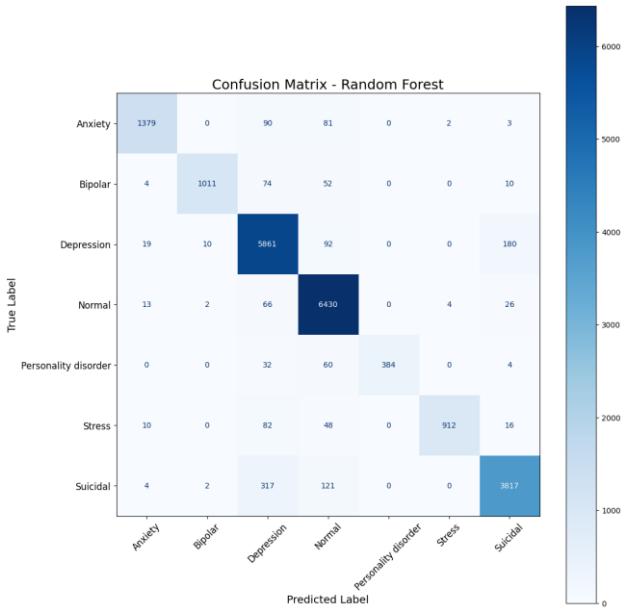

Figure 6. Confusion Matrix Best Model – Random Forest

The confusion matrix in Figure 6 corresponds to the selected best-performing model. High diagonal values confirm strong classification performance, particularly for the normal and depression categories. Misclassifications mainly occur between semantically related classes, such as depression and suicidal, reflecting inherent linguistic overlap in mental health expressions rather than model weakness.

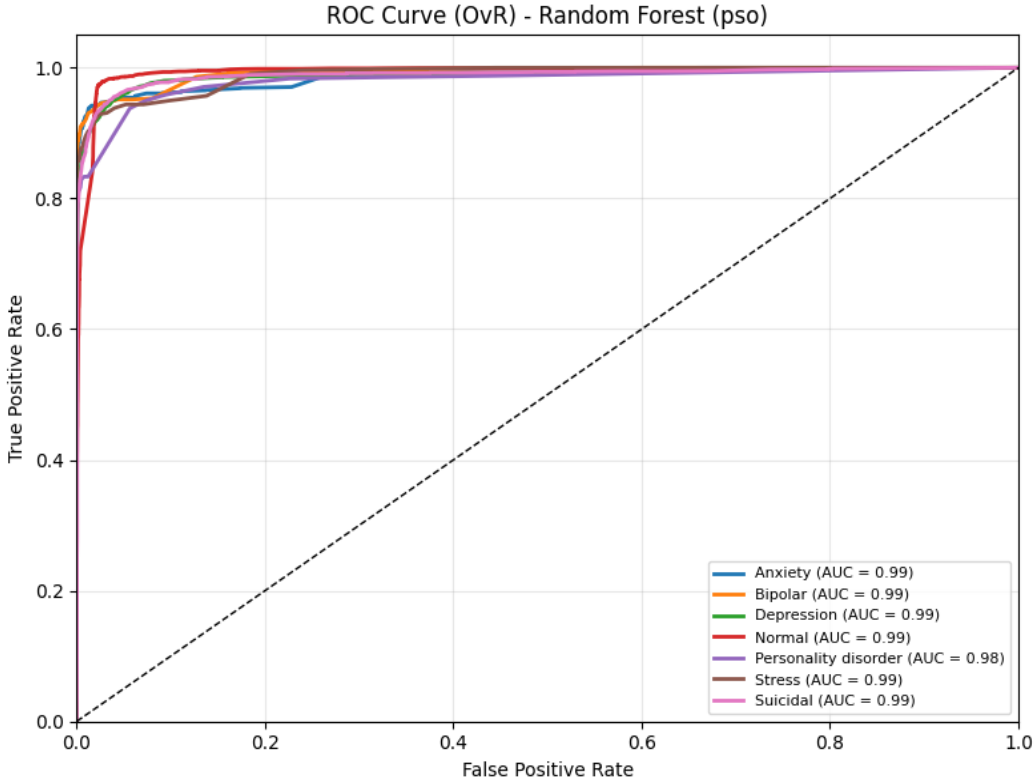

Figure 7. ROC Curve (OvR) – Random Forest (PSO)

Figure 7 presents the ROC curves obtained using a one-vs-rest strategy. AUC values consistently falling between 0.98 and 0.99 across all classes indicate strong discriminative capability. Notably, minority classes also achieve high AUC scores, confirming that the proposed model generalizes well despite class imbalance.

Overall, the experimental findings demonstrate that Particle Swarm Optimization significantly enhances the performance of ensemble-based classifiers. The combination of Random Forest and PSO provides superior accuracy, improved class-level performance, and robust generalization compared to conventional tuning strategies [44], [45], as shown at Table 5.

Table 5. Sample Output of the PSO-Optimized Random Forest Model for Seven-Class Mental Health Status Classification.

Original text	True Label	Predicted Label	Correct	Top Probabilities
morning everyone serious need decent coffee catering open yet	Normal	Normal	✓	Normal (0.954), Depression (0.030), Stres (0.007)
always get fight friend never initiate either everyone always problem idk like everybody get close hate personality wait find right person mask friend anymore fought every single one moved make feel know wrong literally personality wrong ig fight friend	Depression	Depression	✓	Depression (0.855), Suicidal (0.084), Stress (0.038)
fear going doctor finally going doctor really confronting fear swelling pain tonsil bit throat mildly freaking wondering doctor say maybe prescribe something fear hospital crazy doctor say basically ok wo believe hoping basically cure health anxiety know lot people ha believe doctor hope trust enough	Anxiety	Anxiety	✓	Anxiety (0.855), Depression (0.061), Normal (0.023)
due leaving hairbrush home spring break able brush hair month also almost deodorant toothpaste mouth rinse lease apartment running meaning going back home houston get different job actually call work leave cent account wish kidding need clean apartment inspection also need cleaning supply	Stress	Stress	✓	Stress (0.702), Normal (0.160), Suicidal (0.061)

6. CONCLUSION

This study evaluates Particle Swarm Optimization (PSO) as an effective hyperparameter tuning strategy for multi-class mental health status classification using ensemble-based machine learning models. Model performance was systematically assessed by comparing baseline configurations with conventional optimization techniques (Grid Search and Random Search) and metaheuristic optimization across multiple supervised learning algorithms. Experimental results demonstrate that ensemble-based models consistently outperform single classifiers across all evaluation metrics. In particular, the Random Forest model optimized via PSO achieved the highest numerical performance, yielding an accuracy of 0.9329, a macro F1-score of 0.923, and a ROC AUC of 0.989. These findings confirm that PSO enhances model generalization and improves class-balanced performance in imbalanced multi-class mental health datasets.

When positioned against recent literature, the proposed framework demonstrates several distinctive advantages. Compared with binary or limited-class ensemble approaches, such as the fastText–XGBoost model proposed by Ghosal and Jain, the present study extends classification toward seven coexisting mental health categories, improving practical applicability in real-world screening scenarios. In contrast to hierarchical and deep learning–driven frameworks, including the multi-level classification approach by Sutranggono et al. and the hybrid opinion-enhanced architecture proposed by Hossain et al., this work achieves competitive performance while maintaining substantially lower computational complexity, improved reproducibility, and easier deployment. Furthermore, while benchmarking-oriented studies comparing traditional NLP pipelines and large language models, such as Kallstenius et al., highlight the competitiveness of classical feature engineering, they do not explicitly investigate systematic ensemble optimization strategies. Unlike the Genetic Algorithm–based optimization employed by Juanita et al. for three-class clinical classification, this study adopts a swarm-based optimization paradigm and scales the task to a broader seven-class setting under a unified and reproducible experimental framework.

Collectively, these results confirm the three principal contributions of this study: (1) the development of a seven-class mental health classification pipeline using lightweight ensemble-based machine learning with enhanced linguistic preprocessing and data augmentation, (2) the effective application of Particle Swarm Optimization for Random Forest hyperparameter tuning to improve class-balanced generalization, (3) the establishment of a computationally efficient and reproducible comparative evaluation framework suitable for scalable deployment, and (4) the successful integration of the trained model into a web-based application that enables real-time prediction, user interaction, and operational decision-support functionality.

From a practical perspective, the proposed approach demonstrates strong discriminative capability across all mental health categories, including minority and high-risk classes, as evidenced by the confusion matrix and ROC

analyses. Misclassifications are primarily observed among semantically overlapping categories, reflecting inherent linguistic ambiguity rather than model limitations. Although the system is not intended to replace clinical diagnosis, it may serve as a complementary decision-support tool for large-scale mental health screening, triage, and digital monitoring.

Future work may investigate algorithm-level or hybrid imbalance learning approaches, such as class weighting, cost-sensitive learning, or focal loss combined with data augmentation, to further improve minority-class recognition. It may also explore the integration of contextual and temporal information, the incorporation of clinically validated datasets, and the application of advanced interpretability and fairness-aware techniques to further enhance the transparency, reliability, and ethical deployment of real-world mental health assessment systems.

Acknowledgments

The authors express their gratitude to Universitas Logistik dan Bisnis Internasional for providing academic facilities and institutional support that contributed to the completion of this study. The authors also thank colleagues and academic mentors for their insightful discussions, guidance, and constructive feedback during the research process. In addition, appreciation is extended to Muhamad Faizan, whose publicly available Kaggle code served as a reference and was further adapted in this work.

REFERENCES

[1] N. K. Iyortsuun, S. Kim, M. Jhon, and H. Yang, **A Review of Machine Learning and Deep Learning Approaches on Mental Health Diagnosis**, *Healthcare*, Vol. 11, No. 3, pp. 1–27, 2023, https://doi.org/10.3390/healthcare11030285.

[2] U. Madububambachu, A. Ukpebor, and U. Ihezue, **Machine Learning Techniques to Predict Mental Health Diagnoses : A Systematic Literature Review**, *Clin Pract Epidemiol Ment Health*. no. Ml, pp. 1–16, 2024, https://doi.org/10.2174/0117450179315688240607052117.

[3] M. Garg, **Mental Health Analysis in Social Media Posts: A Survey**, *Archives of Computational Methods in Engineering*, vol. 30, no. 3, pp. 1819–1842, 2023, https://doi.org/10.1007/s11831-022-09863-z.

[4] J. Kim, D. Lee, and E. Park, **Machine Learning for Mental Health in Social Media**. J Med Internet Res, vol. 23, pp. 1–17, 2021, https://doi.org/10.2196/24870.

[5] G. Fu *et al.*, **Distant supervision for mental health management in social media: Suicide risk classification system development study**, *J. Med. Internet Res.*, vol. 23, no. 8, 2021, https://doi.org/10.2196/26119.

[6] A. Le Glaz, Y. Haralambous, P. Lenca, and R. Billot, **Machine Learning and Natural Language Processing in Mental Health**, *J. Med. Internet Res*, vol. 23, https://doi.org/10.2196/15708.

[7] Palanivinayagam, C. Z. El-Bayeh, and R. Damaševičius, Twenty Years of Machine-Learning-Based Text Classification: A Systematic Review, Algorithms, vol. 16, no. 5, pp. 1–28, 2023, https://doi.org/10.3390/a16050236.

[8] A.G. Putrada, N. Alamsyah, S. F. Pane, and M. Nurkamal Fauzan, **Feature Importance on Text Analysis for a Novel Indonesian Movie Recommender System,** *in 2023 11th International Conference on Information and Communication Technology (ICoICT),* 2023, pp. 34–39. https://doi.org/10.1109/ICoICT58202.2023.10262504.

[9] O. D. Adeniji, S. O. Adeyemi, and S. A. Ajagbe, **An Improved Bagging Ensemble in Predicting Mental Disorder Using Hybridized Random Forest - Artificial Neural Network Model,** *Informatica*, vol. 46, pp. 543–550, 2022, https://doi.org/10.31449/inf.v46i4.3916.

[10] Konda Vaishnavi, U Nikhitha Kamath, Ashwath Rao B, N V Subba Reddy, **Predicting Mental Health Illness using Machine Learning Algorithms**, *J. of Physics: Conference Series*, Vol. 2161, *1st International Conference on Artificial Intelligence, Computational Electronics and Communication System (AICECS 2021)* 28-30 October 20212022, https://doi.org/10.1088/1742-6596/2161/1/012021

[11] Q. Li and J. Zhou, **A Comparative Analysis of Extreme Gradient Boosting, Decision Tree, Support Vector Machines, and Random Forest Algorithm in Data Analysis of College Students' Psychological Health**, *Informatica*, vol. 49, No. 15, pp. 127–134, 2025, doi: https://doi.org/10.31449/inf.v49i15.7004.

[12] Y. Xin and X. Ren, **Predicting depression among rural and urban disabled elderly in China using a random forest classifier**, *BMC Psychiatry*, pp. 1–11, 2022, https://doi.org 10.1186/s12888-022-03742-4.

[13] B. Ramadhan and S. F. Pane, **Pengaruh Hyperparameter Tuning untuk Efektivitas pada Pendekatan Hybrid dalam Mendiagnosis Stres dan Depresi: Tinjauan Studi Literatur**, *Jurnal Tekno Insentif*, vol. 18, no. 2, pp. 104–118, 2024, https://doi.org/10.36787/jti.v18i2.1516.

[14] R. Choirunnisa, M. Anshori, and W. T. Kusuma**, Improving Random Forest Evaluation in Mental Health Disorder Identification with Cross Validation**, *RIGGS Journal of Artificial Intelligence and Digital Business*, vol. 4, no. 2, pp. 3526–3534, 2025, https://doi.org/10.31004/riggs.v4i2.1053.

[15] M. Jain, V. Saihjpal, N. Singh, and S. B. Singh, **An Overview of Variants and Advancements of PSO Algorithm**, *Applied Sciences* (Switzerland), vol. 12, no. 17, pp. 1–21, 2022, doi: 10.3390/app12178392.

[16] D.-D. Ramírez-Ochoa, L. A. Pérez-Domínguez, E.-A. Martínez-Gómez, and D. Luviano-Cruz, **PSO, a Swarm Intelligence-Based Evolutionary Algorithm as a Decision-Making Strategy: A Review**, Symmetry, Vol. 14, No. 3, 2022, https://doi.org/10.3390/sym14030455

[17] J. Chung and J. Teo, **Mental Health Prediction Using Machine Learning: Taxonomy, Applications, and Challenges**, *Applied Computational Intelligence and Soft Computing,* vol. 2022, 2022, https://doi.org/10.1155/2022/9970363.
[18] N. Dhariwal et al., **A pilot study on AI-driven approaches for classification of mental health disorders,** *Front. Hum. Neurosci*., vol. 18, 2024, https://doi.org/10.3389/fnhum.2024.1376338.
[19] T. Kallstenius, A. J. Capusan, G. Andersson, and A. Williamson, **Comparing traditional natural language processing and large language models for mental health status classification: a multi-model evaluation**, *Sci. Rep.,* vol. 15, no. 1, pp. 1–13, 2025, https://doi.org/10.1038/s41598-025-08031-0.
[20] T. Xiao, D. Ren, S. Lei, J. Zhang, and X. Liu, **Based on grid-search and PSO parameter optimization for Support Vector Machine**, *in Proceeding of the 11th World Congress on Intelligent Control and Automation*, 2014, pp. 1529–1533. https://doi.org/10.1109/WCICA.2014.7052946.
[21] K. Shiomi, T. Sato, and E. Kita, **Comparison of Particle Swarm Optimization Algorithms in Hyperparameter Optimization Problem of Multi Layered Perceptron,** *Computer Assisted Methods in Engineering and Science,* vol. 32, no. 1, pp. 3–24, 2025, https://doi.org/10.24423/cames.2025.1730.
[22] H. S. Salem, M. A. Mead, and G. S. El-Taweel, **Particle Swarm Optimization-Based Hyperparameters Tuning of Machine Learning Models for Big COVID-19 Data Analysis**, *Journal of Computer and Communications*, vol. 12, no. 03, pp. 160–183, 2024, https://doi.org/10.4236/jcc.2024.123010.
[23] S. Ghosal and A. Jain, **Depression and Suicide Risk Detection on Social Media using fastText Embedding and XGBoost Classifier**, *Procedia Comput. Sci.,* vol. 218, pp. 1631–1639, 2022, https://doi.org/10.1016/j.procs.2023.01.141.
[24] M. M. Hossain, M. S. Hossain, M. F. Mridha, M. Safran, and S. Alfarhood, **Multi task opinion enhanced hybrid BERT model for mental health analysis**, *Sci. Rep*., vol. 15, no. 1, pp. 1–20, 2025, https://doi.org/10.1038/s41598-025-86124-6.
[25] S. F. Pane, F. Abdullah, and R. Habibi, **Deteksi Emosi Pada Teks Berbahasa Indonesia Menggunakan Pendekatan Ensemble**, *JTT (Jurnal Teknologi Terapan)*, vol. 10, no. 2, p. 80, 2024, https://doi.org/10.31884/jtt.v10i2.551.
[26] A. N. Sutranggono, R. Sarno, and I. Ghozali, **Multi-Class Multi-Level Classification of Mental Health Disorders Based on Textual Data from Social Media,** *The Journal of Information and Communication Technology*, vol. 1, no. v. 1, 1, pp. 117–148, 2023, [Online]. Available: https://books.google.co.id/books?id=5ySg3D3s-sIC
[27] S. Juanita, A. H. Daeli, M. Syafrullah, W. Anggraeni, and M. H. Purnomo, **A machine learning approach for text pattern diagnosis in mental**

health consultations, *Decision Analytics Journal*, vol. 15, no. March, p. 100572, 2025, https://doi.org/10.1016/j.dajour.2025.100572.

[28] T. Zhang, A. M. Schoene, S. Ji, and S. Ananiadou, **Natural language processing applied to mental illness detection: a narrative review**, *NPJ Digit. Med.*, vol. 5, no. 1, pp. 1–13, 2022, https://doi.org/10.1038/s41746-022-00589-7.

[29] H. T. Duong and T. A. Nguyen-Thi, **A review: preprocessing techniques and data augmentation for sentiment analysis**, *Comput. Soc. Netw.*, vol. 8, no. 1, pp. 1–16, 2021, https://doi.org/10.1186/s40649-020-00080-x.

[30] M. A. Palomino and F. Aider, **Evaluating the Effectiveness of Text Pre-Processing in Sentiment Analysis**, *Applied Sciences (Switzerland)*, vol. 12, no. 17, 2022, https://doi.org/10.3390/app12178765.

[31] M. N. Fauzan, A. G. Putrada, N. Alamsyah, and S. F. Pane, **PCA-AdaBoost Method for a Low Bias and Low Dimension Toxic Comment Classification**, in *2022 International Conference on Advanced Creative Networks and Intelligent Systems (ICACNIS)*, 2022, pp. 1–6. https://doi.org/10.1109/ICACNIS57039.2022.10055017.

[32] D. A. Anggoro and S. S. Mukti, **Performance Comparison of Grid Search and Random Search Methods for Hyperparameter Tuning in Extreme Gradient Boosting Algorithm to Predict Chronic Kidney Failure**, *International Journal of Intelligent Engineering and Systems*, vol. 14, no. 6, pp. 198–207, 2021, https://doi.org/10.22266/ijies2021.1231.19.

[33] L. Abualigah, **Particle Swarm Optimization: Advances, Applications, and Experimental Insights**, *Computers, Materials and Continua*, vol. 82, no. 2, pp. 1539–1592, 2025, https://doi.org/10.32604/cmc.2025.060765.

[34] J. Veintimilla-Reyes, B. Guerrero, D. Maldonado-Segarra, and R. Ortíz-Gaona, **Application of a Heuristic Model (PSO—Particle Swarm Optimization) for Optimizing Surface Water Allocation in the Machángara River Basin, Ecuador**, *Water (Switzerland)*, vol. 17, no. 16, pp. 1–26, 2025, https://doi.org/10.3390/w17162481.

[35] N. Maldini *et al.*, "Deteksi Dini Gangguan Kesehatan Mental berdasarkan Sentimen menggunakan Metode Stacking Early Detection of Mental Health Disorders based on Sentiment using Stacking Method," *Sistemasi: Jurnal Sistem Informasi*, vol. 14, no. 1, pp. 2540–9719, 2025, doi: https://doi.org/10.32520/stmsi.v14i1.4842.

[36] S. Basuki, R. Indrabayu, and N. A. Effendy, **Automatic Categorization of Mental Health Frame in Indonesian X (Twitter) Text using Classification and Topic Detection Techniques,** *Khazanah Informatika : Jurnal Ilmu Komputer dan Informatika*, vol. 10, no. 2, pp. 104–110, 2025, https://doi.org/10.23917/khif.v10i2.3328.

[37] C. Sintiya, G. H. Hutagaol, D. Bate`e, and S. Irviantina, **Evaluasi Teknik Resampling untuk Class Balancing dalam Analisis Sentimen**

Kesehatan Mental Berbasis Bi-LSTM, *Jurnal Sifo Mikroskil*, vol. 26, no. 2, pp. 257–274, 2025, https://doi.org/10.55601/jsm.v26i2.1799.

[38] I. Villanueva-Miranda, Y. Xie, and G. Xiao, **Sentiment analysis in public health: a systematic review of the current state, challenges, and future directions,** *Front. Public Health,* vol. 13, 2025, https://doi.org/10.3389/fpubh.2025.1609749.

[39] M. R. Alfreda Cecio Salwa Alexita, Pramesti Kusumaningtyas, **Optimasi Algoritma Random Forest Menggunakan PSO Untuk Klasifikasi Kanker Payudara Dengan Citra Mammograms**, *Teknika STTKD Jurnal Teknik Elektronik Engine*, vol. 11, no. 1, pp. 47–54, 2025, doi: https://doi.org/10.56521/teknika.v11i1.1346.

[40] N. Agrawal, M. Mahrishi, M. K. Gupta, and M. K. Bohra, **Large scale multi-class pest image classification using structurally adapted DenseNet architecture**, *Sci. Rep.,* vol. 16, no. 1, Dec. 2026, https://doi.org/10.1038/s41598-026-49685-8.

[41] M. Salmi, D. Atif, D. Oliva, A. Abraham, and S. Ventura, **Handling imbalanced medical datasets: review of a decade of research**, *Artif. Intell. Rev*., vol. 57, no. 10, Oct. 2024, https://doi.org/10.1007/s10462-024-10884-2.

[42] D. Remawati, E. Noersasongko, A. Marjuni, and Pujiono, "Mental Health Detection with TF-IDF Feature Extraction," in 2024 IEEE International Conference on Artificial Intelligence and Mechatronics Systems (AIMS), 2024, pp. 1–6. doi: 10.1109/AIMS61812.2024.10512480.

[43] J. Chung and J. Teo, **Single classifier vs. ensemble machine learning approaches for mental health prediction**, *Brain Inform*., vol. 10, no. 1, pp. 1–10, 2023, https://doi.org/10.1186/s40708-022-00180-6.

[44] T. Zhang, K. Yang, S. Ji, and S. Ananiadou, "Emotion fusion for mental illness detection from social media: A survey," Information Fusion, vol. 92, no. December 2022, pp. 231–246, 2023, doi: 10.1016/j.inffus.2022.11.031.

[45] A. V. Lakra and S. Jena, **Optimization of Random Forest Hyperparameter Using Improved PSO for Handwritten Digits Classification**, In: Panda, S.K., Rout, R.R., Sadam, R.C., Rayanoothala, B.V.S., Li, KC., Buyya, R. (eds) *Computing, Communication and Learning*. CoCoLe 2022. *Communications in Computer and Information Science*, vol 1729. Springer, Cham. https://doi.org/10.1007/978-3-031-21750-0_23S.

[46] A. Salloum, A. Almomani, K. M. Alomari, T. Khdour, and M. Alauthman, **Predicting Thyroid Dysfunction Using Classical Machine Learning with Rigorous Statistical Evaluation**, *IEEE Access*, 2026, https://doi.org/10.1109/ACCESS.2026.3666210.