Received: May 8, 2024. Revised: June 21, 2024.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Multi Level Learning based Hierarchical Convolutional Neural Network-Long Short-Term Memory Model for Depression Intensity Prediction

Suganthi Velu[1]* **Punithavalli Muthuswamy[1]**

[1]*Department of Computer Applications, Bharathiar University, Coimbatore - 641046, Tamil Nadu, India*
* Corresponding author's Email: v.suganthiphd@gmail.com

Abstract: Predicting the intensity of a user's depression accurately is crucial for early intervention and preventing mental and physical health issues. Social media platforms are valuable sources of data for understanding and addressing mental health concerns. From this viewpoint, a Multi-Feature Long Short-Term Memory (MF-LSTM) network was developed to estimate depression levels in social media users based on various features: user-specific, community-specific, emotional, and structural features. However, the fixed values of hyperparameters used in MF-LSTM lead to overfitting issues and degrade accuracy. To address this, the MF-BERT-Optimized LSTM (MF-BERT-OLSTM) model is proposed in this article. This model uses an Improved Osprey Optimization Algorithm (IO2A) to optimize the LSTM network hyperparameters. The IO2A mimics the hunting behavior of ospreys for fish from the sea, focusing on balancing exploration and exploitation search processes. The analogy of osprey's hunting behavior is used to find hyperparameters of LSTM in this paper. This IO2A adopts the Circle chaotic mapping in population initialization to improve diversity among individuals and quality of initial solutions. It also uses a dynamically modifiable elite leadership strategy for location updating, balancing global search capabilities and convergence speed. Additionally, a dynamic chaotic weight factor is employed to improve local search capabilities. Thus, the LSTM network parameters are optimized for training and predicting users' depression levels. Additionally, this model uses BERT to better encode the complete semantic information of the given tweet sequences. The experimental results show that the MF-BERT-OLSTM model outperforms the MF-LSTM, Mood2Content, Ensemble, and MentalBERT models on the COVID-19 Twitter Dataset and Post-Traumatic Twitter Dataset. It achieves an accuracy of 98.9% and 98.67%, as well as Root Mean Square Error (RMSE) values of 0.1846 and 0.1157, respectively, on the COVID-19 and Post-Traumatic Twitter Datasets.

Keywords: Mental health, Social media platforms, Depression intensity prediction, MF- LSTM, BERT, Hyperparameter selection, Osprey optimization.

1. Introduction

Stress-related mental illness is a leading cause of global suffering and mortality, particularly among young people, affecting around 280 million individuals and rising daily [1]. Stress is the second most common factor in youth deaths, leading to social isolation and career setbacks [2]. India has the highest rates of depression, psychosis, and bipolar disorder worldwide [3]. Despite available therapies, only 10% of patients receive counseling due to societal stigma, preventing over 70% from seeking help and worsening their mental health [4, 5]. Recent events like movement restrictions, job losses, unemployment, domestic violence, and cyberattacks have exacerbated stress levels, causing dissatisfaction, worry, discomfort, and increased stress among individuals [6, 7]. Research on the mental health of social media users has been limited by time, cost, and lack of information. Online behavioral tests can help identify mental health issues, such as anxiety, which can vary in severity [8]. Treatment decisions are based on symptom severity, and medical intervention may be ineffective if the patient also shows signs of depression or anxiety. To address these challenges, Ghosh et al. [9] developed a deep classifier using webpage data to assess stress levels. They extracted psychological, social, cognitive, individual-level, and distress-dependent n-gram features to profile each user. This information was then input into a small LSTM model with Swish as an activation function. However, existing social media analytics tools often overlook the structural properties of influential entities. To overcome this limitation, the MF-LSTM network model was introduced [10] to estimate stress severity by considering the web topology of influential groups, impact metrics, and various characteristics. It analyzes the structural features of these groups, such as clinicians, academics, and social activists, to evaluate their local and global influence.

1.1 Problem statement

The accuracy of MF-LSTM for predicting depression intensity can be affected by improper hyperparameter selection. Manually determining the best hyperparameter values for different dataset sizes involves numerous trial and error attempts, making it a challenging task. So, an automated and adaptive hyperparameter selection algorithm is needed to improve accuracy, especially with large datasets.

1.2 Major contributions of the manuscript

This manuscript introduces the MF-BERT-OLSTM model to optimize hyperparameters and enhance depression intensity estimation accuracy. This model incorporates the IO2A for optimal hyperparameter selection, which is inspired by the hunting behavior of ospreys. This study introduces three enhancements to the classical O2A algorithm: (i) The use of Circle chaotic mapping for population initialization improvement. (ii) A dynamic elite leadership strategy with a modifiable ratio for location updating. (iii) A dynamic chaotic weight factor for enhancing local searchability. Following this approach, the optimal hyperparameters of the LSTM network are selected for model training and depression intensity prediction. Additionally, this model utilizes a BERT-based encoder to encode the complete semantic information of the given tweet sequence, thereby enhancing prediction accuracy.

The following sections are organized as follows: Section 2 covers the literature survey. Section 3 describes the MF-BERT-OLSTM model and Section 4 demonstrates its test results. Section 5 précises the study and offers future enhancements.

2. Literature survey

Recent studies have shown that DLM are more accurate than traditional machine learning methods in predicting depression levels in social media users. This section will review recent research on using DLMs to identify depression from social media data.

2.1 Depression identification on social media

A hybrid DLM [11] was created to predict a user's mental state by classifying depressive and non-depressive tweets using a grouping of Convolutional Neural Network (CNN) and Bidirectional LSTM (BiLSTM). A Sentic GCN, a Graph Convolutional Network (GCN) based on SenticNet [12], has been created to enhance text dependency graphs by integrating affective knowledge from SenticNet.

A hybrid Sequence, Semantic, Context Learning (SSCL) framework [13] was developed by merging GloVe for feature extraction, LSTM and CNN for capturing the sequence and semantics of tweets, and GRUs with a self-attention strategy for contextual and inherent features. A fully connected layer with a sigmoid function was utilized to detect depression.

The Multi-Aspect Depression Detection with Hierarchical Attention Network (MDHAN) [14] was created to identify depressed users on social media. Small deep-transfer learning language models [15] have been developed to classify depression from tweets.

Ensemble hybrid learning [16] was developed that combines Attention LSTM and Logistic Regression (LR) for automated depression detection. A new typology [17] was created to diagnose depression severity from social media texts. The BiLSTM network with a DistilBERT model was utilized for classification.

A Multimodal Hierarchical Attention (MHA) model [18] was created to detect depression in social media using an attention mechanism. An attention-based BiLSTM-CNN [19] framework was developed to identify depressive tweets.

A new method for detecting depression in social media text [20] was introduced by incorporating linguistic information into transformer models like BERT and MentalBERT. A new hybrid DLM using Fasttext CNN with LSTM [21], was developed for text representations.

2.2 Depression detection during COVID-19

A neural network model was presented [22] to predict post-traumatic stress disorder in adults. The pre-trained BERT and neural network [23] was developed to predict the sentiments from the tweets shared during COVID-19. The DNN [24] was presented to predict the depression of the elderly community according to the social factors related to stress, mental conditions, daily changes, and physical

Table 1. Summary of existing works

Ref. No.	Disadvantages
[11]	Using the BERT model was challenging because of the limited sentence length and low accuracy in learning complex features from the input sequence.
[12]	Increasing the number of GCN layers resulted in low accuracy and F1-score.
[13]	The model did not account for depression severity scores, resulting in low accuracy and F1-score for ternary labeled data.
[14]	The recall and accuracy were low because there was insufficient contextual knowledge of the short and long user-generated content.
[15]	The precision and recall were low because the model was trained on short tweets with limited vocabulary. This may make it less effective at predicting depression intensity in longer text passages.
[16]	Inadequate classification performance due to absence of transformer-based pre-trained language model like BERT and semantic features.
[17]	The model's accuracy was low due to the exclusion of many features during training and the lack of data pre-processing.
[18]	The model's accuracy and F1 score are low due to the lack of clear indicators of depression in the data. This makes it difficult for the model to accurately predict depressive tendencies in users.
[19]	The model's accuracy was low due to the presence of words commonly found in non-depressive samples, leading to confusion and misclassification of depressive samples as non-depressive.
[20]	The authors were unable to optimize hyperparameters due to limited access to GPU resources, leading to poor detection performance.
[21]	The fixed hyperparameters may lead to low recall and accuracy values.
[22]	The accuracy of identifying PTSD-positive cases was low due to the limited number of positive samples.
[23]	Low recall and precision were observed due to limitations in collecting the number of available tweets.
[24]	Choosing the kernel width and number of features for a local model can be challenging, resulting in low accuracy.
[25]	The prediction accuracy and RMSE were not optimal due to the small dataset size.
[26]	The hyperparameters were not optimized, which tends to poor precision.
[27]	The accuracy was lower owing to the limited dataset size and potential information loss from encoding only the first 256 tokens of daily tweets.

distancing. Also, the DNN parameters were optimized by the grid search scheme. A stress prediction technique based on the Elastic-Net normalization model [25] was developed to identify depression risks.

A new hybrid DLM [26] using the LSTM and CNN was developed to estimate the consequence of COVID-19 on an individual's psychological strength from tweets. A new Mood2Content model [27] was created to detect depression early in online users. It utilized content and mood encoders to extract content and mood representations from daily tweets. These representations are aggregated and inputted into a user encoder with Transformer and self-attention layers to predict depression risk. Table 1 outlines the drawbacks of above-studies techniques individually.

2.3 Research gap

Existing studies often struggle with selecting optimal hyperparameters, leading to low accuracy in depression detection. This study introduces a new metaheuristic optimization algorithm to dynamically choose hyperparameters for the LSTM network, improving its ability to predict depression severity with higher accuracy. In addition, semantic features of sentences are considered through BERT.

3. Proposed methodology

This section explains the MF-BERT-OLSTM network model for depression intensity prediction in detail. Table 2 lists the notations used in this study.

A conceptual design of this work is portrayed in Fig. 1. Different phases in the depression intensity

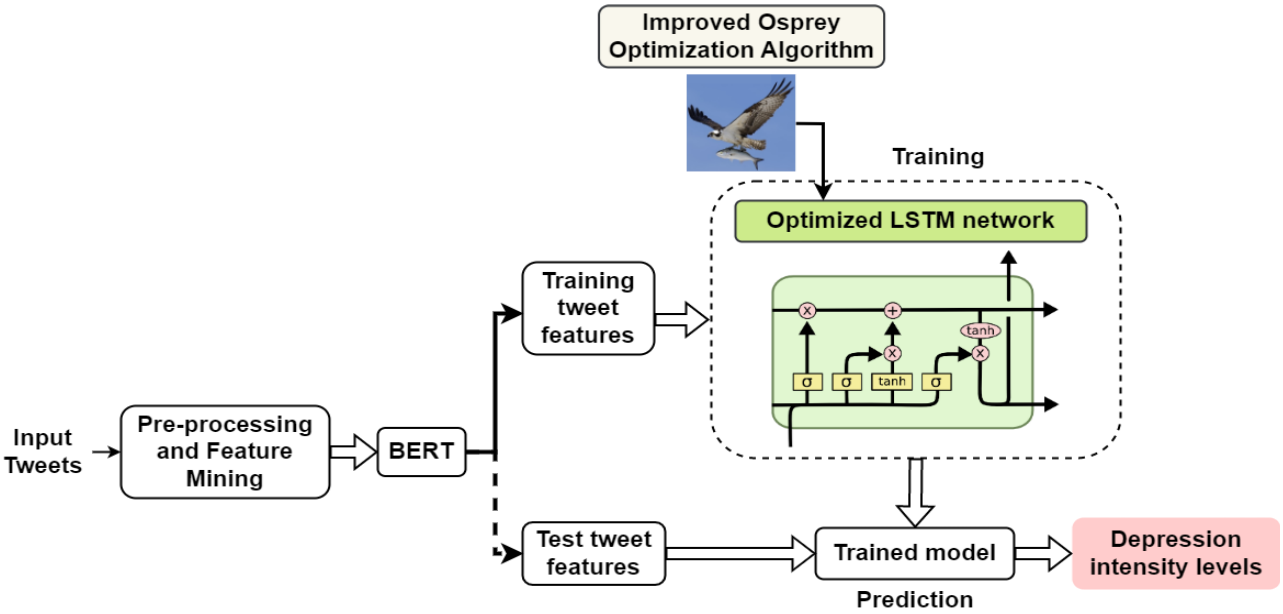

Figure. 1 Conceptual design of the presented work

Table 2. Lists of notations

Notations	Description
P	Population matrix of osprey's positions
P_i and $p_{i,j}$	i^{th} osprey and its j^{th} dimension, respectively
N	Amount of ospreys
m	Amount of problem variables
F	Vector of the fitness values
F_i	Fitness value of the i^{th} osprey
S_i	Group of fish locations for the i^{th} osprey
P_{best}	Optimal candidate solution
$P_i^{S_1}$	New location of the i^{th} osprey according to the initial stage of IO2A
$p_{i,j}^{S_1}$	j^{th} dimension of $P_i^{S_1}$
$F_i^{S_1}$	Fitness value of $P_i^{S_1}$
CS_i	Selected fish for i^{th} osprey
$CS_{i,j}$	j^{th} dimension for CS_i
$I_{i,j}$	Random value in the range of {1,2}
r_{ij}	Random number between [0,1]
$P_{i,j}^{best}$	Position of the individual with optimal fitness value in the population
α	Dynamic modification factor
t	Present iteration
T	Highest amount of iteration
ω	Dynamic chaotic weight factor
$P_i^{S_2}$	New location of the i^{th} osprey according to the second stage of IO2A
$p_{i,j}^{S_2}$	j^{th} dimension of $P_i^{S_2}$
$F_i^{S_2}$	Fitness value of $P_i^{S_2}$

prediction model include (i) data collection, (ii) pre-processing, (iii) feature mining and (iv) depression intensity prediction. First, the dataset is created by collecting user data from Twitter. A depression score is calculated based on sentiment polarities and Latent Semantic Analysis (LSA). This score is classified into no, minor, considerable and severe depression levels. The created dataset is pre-processed and used to extract a variety of features [10]. This involves removing emojis, punctuation, articles, and special characters from the tweets. The content of the tweets is tokenized, stemmed, and lemmatized. Redundant or unwanted words like typographical errors or acronyms are eliminated.

After preprocessing, the online behaviors of Twitter users are considered to extract emotion features, event features, user-specific features and depression-related n-gram features.

Also, social network structural properties and influence features are extracted. The obtained features are passed to the BERT to obtain the global information of the tweets. The BERT consists of many transformer units with a multi-head attention strategy [28]. Input vectors are linearly transformed through multiple layers and then processed by the attention unit to determine attention weights. The attention strategy result is merged with the preceding linear transformation to produce the absolute result of the multi-head attention strategy. The output of the BERT is fed to the LSTM network for depression severity prediction, with model hyperparameters optimized by the IO2A to increase prediction accuracy. The hyperparameters of the LSTM model include word embedding size, number of filters, activation function, learning rate, dropout rate, weight decay, epochs, batch size, momentum rate, and loss function. The values of these hyperparameters are optimally set by the IO2A for effective model training. A detailed description of this IO2A is presented in the section below.

3.1 Improved osprey optimization algorithm for hyperparameter selection

The nocturnal bird of prey known as the osprey can serve as inspiration for a novel optimization algorithm by providing a natural test case in its ability to capture fish and fly them to a more favorable location [29]. The improvements in each phase of the IO2A algorithm are described below:

1. Initialization: In the traditional OOA algorithm, pseudo-random numbers are used to initialize the population position, leading to low-quality initial population individuals and insufficient traversal of the solution space. By replacing pseudo-random numbers with chaotic sequences for population initialization, the initial population diversity is increased, and the initial solution quality is improved. Chaotic mapping also helps reduce randomness fluctuation in population initialization, increasing the robustness.
2. Exploration phase: The IO2A algorithm updates individual positions through a dynamic elite leadership strategy. Initially, this strategy has a small proportion, allowing for random target detection by individuals to explore the solution space fully. As the algorithm progresses, this strategy plays a larger role, guiding individual position updates towards the optimal solution. This reduces ineffective searches and improves convergence speed.
3. Exploitation phase: The IO2A algorithm incorporates a dynamic chaotic weight factor strategy in the exploitation phase. The weight factor, defined by Cubic chaotic mapping, is dynamically adjusted based on the number of iterations. This strategy enhances the algorithm's local search ability and improves optimization accuracy.

The mathematical modeling of the IO2A is explained below.

3.1.1. Initialization

The IO2A is a population-based method, which utilizes the searchability of the population to obtain an optimal result iteratively. Every osprey in the IO2A population is a potential solution because it uses its location in the search space to make decisions about the problem's variables.

First, the IO2A population is modeled by a matrix as given in Eq. (1), where $i = 1, \dots, N$ and $j = 1, \dots, m$. Then, the location of ospreys in the search space is set by the Circle chaotic mapping as given in Eq. (2). In Eqs. (1) and (2), P is the population matrix of osprey's positions, P_i denotes the i^{th} osprey (i.e., a candidate solution), $p_{i,j}$ indicates its j^{th} dimension (i.e., problem variable), N is the amount of ospreys, and m is the amount of problem variables (i.e., number of hyperparameters of the LSTM model.

$$P = \begin{bmatrix} P_1 \\ \vdots \\ P_i \\ \vdots \\ P_N \end{bmatrix}_{N \times m} = \begin{bmatrix} p_{1,1} & \cdots & p_{1,j} & \cdots & p_{1,m} \\ \vdots & \ddots & \vdots & \iddots & \vdots \\ p_{i,1} & \cdots & p_{i,j} & \cdots & p_{i,m} \\ \vdots & \iddots & \vdots & \ddots & \vdots \\ p_{N,1} & \cdots & p_{N,j} & \cdots & p_{N,m} \end{bmatrix}_{N \times m} \quad (1)$$

$$p_{i,j} = \mathrm{mod}\left(p_i + 0.2 - \left(\frac{0.5}{2\pi}\right)\sin(2\pi p_i), 1\right) \quad (2)$$

The objective (fitness) function of a problem such as prediction accuracy can be evaluated by comparing each osprey as a candidate solution, and the evaluated values can be represented using a vector as given in Eq. (3):

$$F = \begin{bmatrix} F_1 \\ \vdots \\ F_i \\ \vdots \\ F_N \end{bmatrix}_{N \times 1} = \begin{bmatrix} F(X_1) \\ \vdots \\ F(X_i) \\ \vdots \\ F(X_N) \end{bmatrix}_{N \times 1} \quad (3)$$

In Eq. (3), F denotes the vector of the fitness values and F_i denotes the fitness value of the i^{th} osprey. The calculated fitness values are essential for determining the viability of potential solutions. The best value denotes the optimal candidate result, while the worst value denotes the worst solution. Each iteration of the search process involves updating the ospreys' position in the search space, which in turn necessitates updating the best candidate solution.

3.1.2. Exploration – searching location and hunting prey

Ospreys, with their strong eyesight, can detect fish underwater and attack them. The primary stage of population update in IO2A is designed according to this nature. This simulation modifies the osprey's location in the search space, increasing the searchability of IO2A in recognizing best areas and evading local optima. The IO2A design considers underwater fishes with better fitness function values for each osprey's position in the search space. The group of fish for every osprey is signified by Eq. (4).

$$S_i = \{P_k | k \in \{1, \dots, N\} \wedge F_k < F_i\} \cup \{P_{best}\} \quad (4)$$

In Eq. (4), S_i denotes the group of fish locations for the i^{th} osprey and P_{best} denotes the optimal candidate solution (i.e., the best osprey). In this phase, individuals randomly select attack targets (fish) for location updating to explore the search space and avoid local optima. However, this random strategy may not always yield better solutions and can lead to an increase in invalid searches over iterations. To address this, a dynamic elite leadership strategy with a modifiable ratio is introduced. So, a new location for the corresponding osprey is determined as given in Eqs. (5a) and (5b):

$$p_{i,j}^{S_1} = p_{i,j} + \alpha \cdot r_{ij} \cdot \left(P_{i,j}^{best} - P_{i,j}\right) + (1 - \alpha) \cdot r_{ij} \cdot \left(CS_{i,j} - I_{i,j} \cdot p_{i,j}\right), \alpha = \frac{t}{T} \quad (5a)$$

$$p_{i,j}^{S_1} = \begin{cases} p_{i,j}^{S_1}, & lb_j \le p_{i,j}^{S_1} \le ub_j \\ lb_j, & p_{i,j}^{S_1} < lb_j \\ ub_j, & p_{i,j}^{S_1} > ub_j \end{cases} \quad (5b)$$

When the fitness function value increases, this new location changes the former location of the osprey as shown in Eq. (6):

$$P_i = \begin{cases} P_i^{S_1}, & F_i^{S_1} < F_i \\ P_i, & Otherwise \end{cases} \quad (6)$$

In Eqs. (5a), (5b), and (6), $P_i^{S_1}$ defines the new location of the i^{th} osprey according to the initial stage of IO2A, $p_{i,j}^{S_1}$ denotes its j^{th} dimension, $F_i^{S_1}$ indicates its fitness value, CS_i denotes the chosen fish for i^{th} osprey, $CS_{i,j}$ is its j^{th} dimension, $I_{i,j}$ is random value in the range of $\{1,2\}$ and r_{ij} is a random number between $[0,1]$. Also, $P_{i,j}^{best}$ is the position of the individual with optimal fitness value in the population and α is a dynamic modification factor to control the ration between the elite bootstrapping strategy and randomized exploration, whereas α increases linearly from 0 to 1 with the number of iterations.

This IO2A uses α to shift from random exploration to elite leadership as the amount of iteration increases. A smaller value of α at the start promotes exploration and avoids local optima, while a gradually increasing α focuses on elite leadership for faster convergence to the best solution, reducing invalid searches.

3.1.3 Exploitation – Moving the fish to the safe location

In this phase, individuals who find locally optimal solutions undergo a recalibration process to slightly adjust their positions in the search space. This helps the algorithm move away from local optima. However, this method does not fully leverage global optimal solutions and may not provide accurate local search.

To address this, a dynamic chaotic weighting factor is proposed based on Cubic chaotic mapping. This factor adjusts weight coefficients during iterations to enhance local search. The dynamic chaotic weight factor is calculated by Eq. (7):

$$\omega(t+1) = \frac{2.595\omega(t)\left(1-\omega(t)^2 \cdot (T-t)\right)}{T} \quad (7)$$

In Eq. (7), $\omega(1)$ takes the value 0.3, t $(t = 1, \dots, T)$ is the present iteration and T indicates the highest amount of iteration. So, the location of the introduced dynamic chaos, the weight factor is formulated as given by Eqs. (8a) and (8b):

$$p_{i,j}^{S_2} = \omega(t) \cdot p_{i,j} + \frac{lb_j + r_{ij} \cdot (ub_j - lb_j)}{t} \quad (8a)$$

$$p_{i,j}^{S_2} = \begin{cases} p_{i,j}^{S_2}, & lb_j \le p_{i,j}^{S_2} \le ub_j \\ lb_j, & p_{i,j}^{S_2} < lb_j \\ ub_j, & p_{i,j}^{S_2} > ub_j \end{cases} \quad (8b)$$

After that, if the fitness function value is increased in this new location, it changes the former location of the resultant osprey as given in Eq. (9):

$$P_i = \begin{cases} P_i^{S_2}, & F_i^{S_2} < F_i \\ P_i, & Otherwise \end{cases} \quad (9)$$

In Eqs. (8a), (8b), and (9), $P_i^{S_2}$ defines the new location of the i^{th} osprey according to the second stage of IO2A, $p_{i,j}^{S_2}$ denotes its j^{th} dimension and $F_i^{S_2}$ indicates its fitness function value. Introducing the dynamic chaotic weight factor allows for a more precise search near the optimal solution during position updates. The weight factor decreases as the number of iterations increases, transitioning from fine search to rapid convergence.

Received: May 8, 2024. Revised: June 21, 2024.

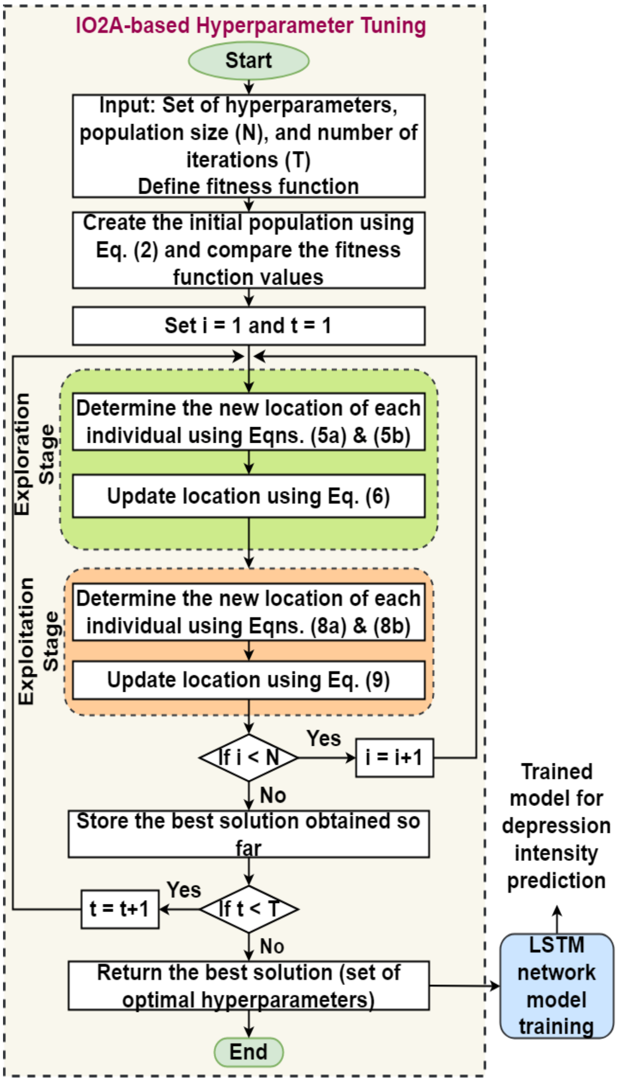

Figure. 2 Flow diagram of IO2A for optimal hyperparameter selection

This approach enhances accuracy in finding the best solution and evading local optima in optimization problems with multiple local optimal values.

Thus, the IO2A updates osprey positions in the first iteration, compares fitness function values, and updates the best candidate solution in the subsequent iterations. The algorithm updates locations for ospreys in the final iteration, and after full implementation, the best candidate solution (best hyperparameters of LSTM network) stored during iterations is considered as a solution to the problem (hyperparameter selection). Algorithm 1 describes the pseudocode of IO2A for hyperparameter selection Fig. 2 depicts its overall workflow.

Algorithm 1: IO2A-based hyperparameter selection

Input: Set of LSTM hyperparameters
Output: Optimal hyperparameters

Begin
//Initialization stage:
Initialize the IO2A population size N, the total number of iterations T and boundary conditions (ub and lb);
Define F (prediction accuracy);
Create the initial population matrix randomly using Eqs. (1) and (2);
$\boldsymbol{while}(t = 1{:}T)$
  $\boldsymbol{for}(i = 1{:}N)$
    Evaluate the fitness function by Eq. (3);
    Specify the target fish population for each individual from Eq. (4);
    //Exploration Stage:
    Determine the new location of each individual using Eq. (5a);
    Verify the boundary criteria by Eq. (5b);
    Update the i^{th} individual using Eq. (6);
    //Exploitation Stage:
    Determine the new location of each individual using Eq. (8a);
    Verify the boundary criteria by Eq. (8b);
    Update the i^{th} individual using Eq. (9);
  $\boldsymbol{end\ for}$
  Update the population optimal fitness value and best location;
  $t = t + 1$;
$\boldsymbol{end\ while}$
Return the best solution (i.e., optimal hyperparameters);
End

Thus, the MF-BERT-OLSTM is trained using a set of optimal hyperparameters and used to accurately predict users' depression levels from their tweets.

4. Results and discussion

In this section, the efficiency of the MF-BERT-OLSTM model is compared to other models including MF-LSTM [10], Ensemble [16], MentalBERT [20] and Mood2Content [27]. The experiments are conducted on a laptop equipped with an Intel® Core ™ i7-1065G7 CPU @ 4.90GHz, 8GB RAM, and a 256GB SSD running Windows 10 64-bit.

To evaluate the performance improvements, all existing and proposed models are implemented in the MATLAB 2019b software with two distinct datasets.

4.1 Parameter settings

Table 3 presents the parameters and their optimal values used to simulate both existing and proposed models, for performance analysis.

Received: May 8, 2024. Revised: June 21, 2024.

Table 3. List of optimal hyperparameters for proposed and existing models

Parameters	Search Space	Optimal Range
Proposed MF-BERT-LSTM, MF-LSTM [10], Mood2Content [27], MentalBERT [20] and Ensemble [16] Models		
Learning rate	[0.0001, 0.1]	0.01
Dropout rate	[2%, 5%]	2%
Weight decay	[0.0001, 0.001]	0.0001
Number of epochs	[100, 400]	250
Batch size	[32, 64, 128, 512]	128
Momentum	[0, 1]	0.9
Optimizer	[Stochastic gradient descent, Adam]	Adam
Loss	[Cross-entropy, mean square error]	Cross-entropy
MF-LSTM [10]		
No. of LSTM layers	[1, 3, 5, 7]	3
No. of hidden units	[12, 24, 48, 96, 128]	128
Word embedding size	[128, 256, 520]	520
Activation function	[linear, ReLU, tan-sigmoid, swish]	Swish
MF-BERT-LSTM		
Max. sequence length	[384, 512]	512
Warmup rate	[0.1, 0.12, 0.2]	0.1
Mood2Content [27]		
Patience of early stop	[10, 20, 50]	10
Modification factor $\boldsymbol{\alpha}$	[0.2, 0.5, 0.9]	0.5
Ensemble Model [16]		
Regularization strength	[0.5, 1]	1
MentalBERT [20]		
Gamma	[0.1, 0.2]	0.1
Step size	[2, 3, 5]	5
Smoothing parameter $\boldsymbol{\alpha}$	[0.001, 0.01]	0.001
Learnable parameter $\boldsymbol{\beta}$	[0.0001, 0.001]	0.0001

Table 4. Dataset details

Dataset	Training Set	Testing Set
COVID-19 Twitter Dataset	8000 non-depression	2000 non-depression
	32000 depression (8000 for each label)	8000 depression (2000 for each label)
Post-Traumatic Twitter Dataset	8000 non-depression	2000 non-depression
	16000 depression and 16000 PTSD	4000 depression and 4000 PTSD

4.2 Dataset description

This study analyzes the following datasets for performance evaluation:

1. COVID-19 Twitter Dataset [10, 27]: The data was gathered from Twitter accounts between March and December 2020, including around 50000 users, with 10000 non-depression and 40000 depression users (10000 for each label). This dataset is compiled by web-crawling Twitter user's profile information and timeline. Additionally, it searches for an anchor tweet that characterizes the user's mental state. For a month, every tweet that the anchor tweet posted was collected. When a user's anchor tweet says "(I'm/was/I am/I've been) declared depression," their tweets are classified as depressed. Individuals who failed to post any tweets that included the terms "stress" or "depressed" were classified as non-depressed. In this way, a non-depression dataset is also prepared. Additionally, an independent re-labeling strategy is developed for classifying the depressed tweets. This strategy uses the LSA to compute a depression score based on the emotion polarity of tweets. The scores are divided into minimal, mild, moderate, and severe depression categories.
2. Post-Traumatic Twitter Dataset: The dataset, named the Post-Traumatic Twitter Dataset, is designed to evaluate the proposed model in this study. This dataset is structured similarly to other datasets like the CLPsych 2015 Shared Task dataset [16] and Dreaddit dataset [20]. It includes user-generated posts from 50000 Twitter users with depression or Post-Traumatic Stress Disorder (PTSD) between January and December 2022. Particularly, there are three labels, i.e., non-depression, depression and PTSD. Of these users, 10000 have no depression, 20000 have depression and 20000 have PTSD. This dataset has been annotated using the Amazon Mechanical Turk. This dataset also includes lexical, syntactic and social media features per post.

Both datasets are divided into a training and a test set with an 80:20 split ratio. See Table 4 for more details.

4.3 Performance evaluation measures

The following metrics are measured to compare the presented and conventional models:

- Accuracy: It is the proportion of accurate identifications made out of all the data tested. It is calculated by Eq. (10).

$$Accuracy = \frac{True\ Positive\ (TP)+True\ Negative\ (TN)}{TP+TN+False\ Positive\ (FP)+False\ Negative\ (FN)} \quad (10)$$

Here, TP is the amount of labels that are identified as no depression, TN is the amount of labels that are identified as depression, FP is the amount of labels wrongly identified as depression and FN is the amount of labels wrongly identified as no depression.

- Precision: It is calculated by Eq. (11).

$$Precision = \frac{TP}{TP+FP} \quad (11)$$

- Recall: It is calculated by Eq. (12).

$$Recall = \frac{TP}{TP+FN} \quad (12)$$

- F1-score: It is measured by Eq. (13).

$$F1 = \frac{2 \times Precision \times Recall}{Precision+Recall} \quad (13)$$

- Root Mean Square Error (RMSE): It defines the disparity of all actual tags (l_a) and estimated tags (l_e). It is calculated by Eq. (14).

$$RMSE = \sqrt{\sum (l_a - l_e)^2} \quad (14)$$

4.4 Performance analysis for COVID-19 twitter dataset

Table 5 displays the confusion matrix of the existing and proposed models using the COVID-19 twitter dataset. Fig. 3 compares various models on the COVID-19 twitter dataset and it is noticed that the MF-BERT - OLSTM outperforms current models in

Table 5. Confusion matrix of existing and proposed models using COVID-19 Twitter Dataset

Models	True Labels						
MentalBERT	Predicted Labels	Labels	1 (No depression)	2 (Minimal depression)	3 (Mild depression)	4 (Moderate depression)	5 (Severe depression)
		1	1785	54	43	40	35
		2	54	1805	41	35	25
		3	62	57	1820	42	45
		4	50	43	43	1840	45
		5	49	41	53	43	1850
Mood2Content		Labels	1	2	3	4	5
		1	1800	50	38	38	35
		2	52	1827	31	33	24
		3	60	52	1850	40	45
		4	46	40	33	1848	44
		5	42	31	48	41	1852
Ensemble		Labels	1	2	3	4	5
		1	1820	45	32	32	32
		2	48	1847	27	30	21
		3	52	45	1866	36	41
		4	42	35	30	1865	41
		5	38	28	45	37	1865
MF-LSTM		Labels	1	2	3	4	5
		1	1840	40	29	28	30
		2	42	1870	25	25	20
		3	48	40	1880	30	38
		4	37	30	26	1887	36
		5	33	20	40	30	1876
Proposed (MF-BERT-OLSTM)		Labels	1	2	3	4	5
		1	1975	6	7	5	6
		2	7	1975	6	6	5
		3	5	6	1974	7	4
		4	6	7	6	1976	5
		5	7	6	7	6	1980

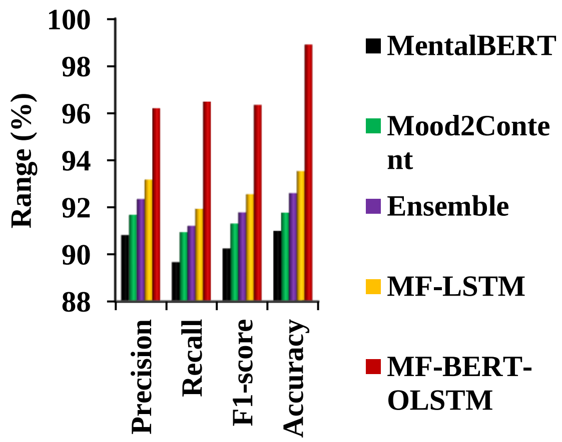

Figure. 3 Comparison of MF-BERT-OLSTM model against existing models on COVID-19 twitter dataset

predicting users' depression severity due to optimized hyperparameters.

Compared to MentalBERT, Mood2Content, ensemble, and MF-LSTM models, MF-BERT-OLSTM increases precision by 5.92%, 4.93%, 4.17%, and 3.25%, recall by 7.59%, 6.09%, 5.78%, and 4.95%, F1-score by 6.75%, 5.51%, 4.97%, and 4.1%, and accuracy by 8.68%, 7.77%, 6.8%, and 5.74%, respectively.

4.5 Performance analysis for post-traumatic twitter dataset

Table 6 displays the confusion matrix of the existing and proposed models using the post-traumatic twitter dataset.

Fig. 4 compares various models on the post-traumatic twitter dataset and it is noted that the MF-BERT-OLSTM outperforms current models in predicting users' depression severity due to optimized hyperparameters and better model training. Compared to all existing models, MF-BERT-OLSTM increases precision by 7.81%, 7.09%, 3.91%, and 2.45%, recall by 7.9%, 7.29%, 3.98%, and 2.5%, F1-score by 7.85%, 7.18%, 3.94%, and 2.47%, and accuracy by 7.72%, 7.01%, 3.87%, and 2.43%, respectively. Fig. 5 compares the RMSE of the proposed and existing models on the two different datasets. The RMSE of the MF-BERT-OLSTM is significantly lower compared to MentalBERT, Mood2Content, ensemble and MF-LSTM models on both datasets.

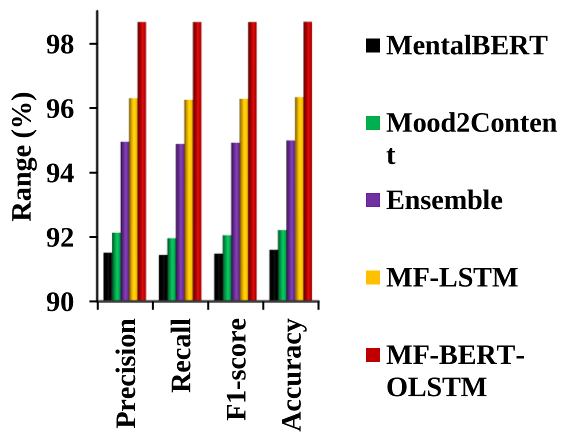

Figure. 4 Comparison of MF-BERT-OLSTM model against existing models on post-traumatic twitter dataset

Table 6. Confusion matrix of MF-BERT-OLSTM model using PTSD dataset

Models		True Labels			
		Labels	1 (No depression)	2 (Depression)	3 (PTSD)
MentalBERT	Predicted Labels	1	1730	110	150
		2	140	3730	150
		3	130	160	3700
Mood2Content		Labels	1	2	3
		1	1750	100	140
		2	130	3750	140
		3	120	150	3720
Ensemble		Labels	1	2	3
		1	1802	60	77
		2	98	3852	78
		3	100	88	3845
MF-LSTM		Labels	1	2	3
		1	1840	40	50
		2	80	3888	45
		3	80	72	3905
Proposed (MF-BERT-OLSTM)		Labels	1	2	3
		1	1920	10	17
		2	40	3979	15
		3	40	11	3968

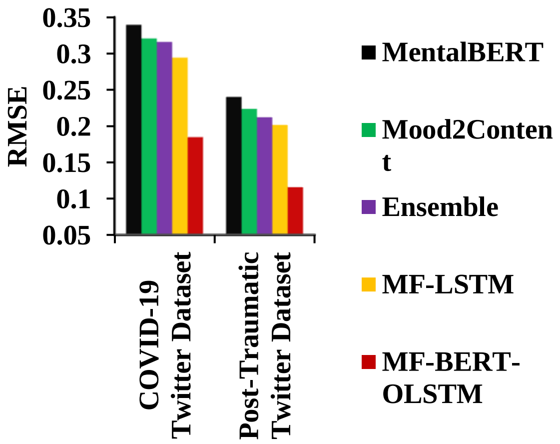

Figure. 5 RMSE for proposed and existing models on two datasets

Specifically, on the COVID-19 dataset, the RMSE reduction ranges from 37.25% to 45.59%, while on the PTSD dataset, the reduction ranges from 42.58% to 51.81%.

5. Conclusion

This study introduces the MF-BERT-OLSTM model, which enhances the prediction performance of depression intensity by optimizing the network's hyperparameters using the IO2A approach. The IO2A is inspired by the hunting behavior of ospreys. By leveraging exploration and exploitation stages, the model identifies the optimal hyperparameters for estimating depression intensity levels of Twitter users. The test results demonstrate that the MF-BERT-OLSTM model achieves 98.9% accuracy and 0.1846 RMSE on the COVID-19 dataset, as well as 98.67% accuracy and 0.1157 RMSE on the PTSD dataset, compared to the MentalBERT, Mood2Content, ensemble, and MF-LSTM models in depression intensity prediction. Future research will explore predicting depression or stress severity on various online networking services like Facebook and Instagram to assess the model's generalizability.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

Conceptualization, methodology, software, validation, Suganthi; formal analysis, investigation, Punithavalli; resources, data curation, writing—original draft preparation, Suganthi; writing—review and editing, Suganthi; visualization, supervision, Punithavalli.

References

[1] Alwakeel, M. Alwakeel, S. R. Zahra, T. J. Saleem, M. Hijji, S. S. Alwakeel, ... and S. Alzorgi, "Common Mental Disorders in Smart City Settings and Use of Multimodal Medical Sensor Fusion to Detect Them", *Diagnostics*, Vol. 13, No. 6, p. 1082, 2023.

[2] S. Frimpong, R. Y. Sunindijo, C. C. Wang, and E. F. Boadu, "Domains of Psychosocial Risk Factors Affecting Young Construction Workers: A Systematic Review", *Buildings*, Vol. 12, No. 3, p. 335, 2022.

[3] World Mental Health Report: Transforming Mental Health for All. Geneva: World Health Organization; 2022. Licence: CC BY-NC-SA 3.0 IGO.

[4] C. F. Reynolds 3rd, D. V. Jeste, P. S. Sachdev, and D. G. Blazer, "Mental Health Care for Older Adults: Recent Advances and New Directions in Clinical Practice and Research", *World Psychiatry*, Vol. 21, No. 3, pp. 336-363, 2022.

[5] E. Mork, S. R. Aminoff, E. A. Barrett, C. Simonsen, W. T. V. Hegelstad, T. V. Lagerberg, ... and K. L. Romm, "COVID-19 Lockdown–Who Cares? The First Lockdown from the Perspective of Relatives of People with Severe Mental Illness", *BMC Public Health*, Vol. 22, No. 1, p. 1104, 2022.

[6] M. R. Miah, M. M. Hasan, J. T. Parisha, and S. H. Chowdhury, "Socioeconomic Impact of the Coronavirus Pandemic with Multiple Factors on Global Healthcare Policy", *Journal of Politics and Law*, Vol. 15, No. 4, pp. 242-283vii, 2022.

[7] E. M. Thompson, L. Destree, L. Albertella, and L. F. Fontenelle, "Internet-Based Acceptance and Commitment Therapy: A Transdiagnostic Systematic Review and Meta-Analysis for Mental Health Outcomes", *Behavior Therapy*, Vol. 52, No. 2, pp. 492-507, 2021.

[8] I. Cataldo, B. Lepri, M. J. Y. Neoh, and G. Esposito, "Social Media Usage and Development of Psychiatric Disorders in Childhood and Adolescence: A Review", *Frontiers in Psychiatry*, Vol. 11, p. 508595, 2021.

[9] S. Ghosh and T. Anwar, "Depression Intensity Estimation via Social Media: A Deep Learning Approach", *IEEE Transactions on Computational Social Systems*, Vol. 8, No. 6, pp. 465-1474, 2021.

[10] V. Suganthi and M. Punithavalli, "User Depression and Severity Level Prediction Using COVID-19 Epidemic from Social Network Data", *ARPN Journal of Engineering and*

Applied Sciences, Vol. 18, No. 10, pp. 1187-1194, 2023.
[11] H. Kour and M. K. Gupta, "An Hybrid Deep Learning Approach for Depression Prediction from User Tweets Using Feature-Rich CNN and Bi-directional LSTM", *Multimedia Tools and Applications*, Vol. 81, No. 17, pp. 23649-23685, 2022.
[12] B. Liang, H. Su, L. Gui, E. Cambria, and R. Xu, "Aspect-Based Sentiment Analysis via Affective Knowledge Enhanced Graph Convolutional Networks", *Knowledge-Based Systems*, Vol. 235, p. 107643, 2022.
[13] A. Nadeem, M. Naveed, M. Islam Satti, H. Afzal, T. Ahmad, and K. I. Kim, "Depression Detection Based on Hybrid Deep Learning SSCL Framework Using Self-Attention Mechanism: An Application to Social Networking Data", *Sensors*, Vol. 22, No. 24, p. 9775, 2022.
[14] H. Zogan, I. Razzak, X. Wang, S. Jameel, and G. Xu, "Explainable Depression Detection with Multi-Aspect Features Using a Hybrid Deep Learning Model on Social Media", *World Wide Web*, Vol. 25, No. 1, pp. 281-304, 2022.
[15] M. Rizwan, M. F. Mushtaq, U. Akram, A. Mehmood, I. Ashraf, and B. Sahelices, "Depression Classification from Tweets Using Small Deep Transfer Learning Language Models", *IEEE Access*, Vol. 10, pp. 129176-129189, 2022.
[16] L. Ansari, S. Ji, Q. Chen, and E. Cambria, "Ensemble Hybrid Learning Methods for Automated Depression Detection", *IEEE Transactions on Computational Social Systems*, Vol. 10, No. 1, pp. 211-219, 2022.
[17] M. Kabir, T. Ahmed, M. B. Hasan, M. T. R. Laskar, T. K. Joarder, H. Mahmud, and K. Hasan, "DEPTWEET: A Typology for Social Media Texts to Detect Depression Severities", *Computers in Human Behavior*, Vol. 139, p. 107503, 2023.
[18] Z. Li, Z. An, W. Cheng, J. Zhou, F. Zheng, and B. Hu, "MHA: A Multimodal Hierarchical Attention Model for Depression Detection in Social Media", *Health Information Science and Systems*, Vol. 11, No. 1, p. 6, 2023.
[19] T. Ghosh, M. H. Al Banna, M. J. Al Nahian, M. N. Uddin, M. S. Kaiser, and M. Mahmud, "An Attention-Based Hybrid Architecture with Explainability for Depressive Social Media Text Detection in Bangla", *Expert Systems with Applications*, Vol. 213, p. 119007, 2023.
[20] L. Ilias, S. Mouzakitis, and D. Askounis, "Calibration of Transformer-Based Models for Identifying Stress and Depression in Social Media", *IEEE Transactions on Computational Social Systems*, Vol. 11, No. 2, pp. 1979-1990, 2023.
[21] V. Tejaswini, K. Sathya Babu, and B. Sahoo, "Depression Detection from Social Media Text Analysis Using Natural Language Processing Techniques and Hybrid Deep Learning Model", *ACM Transactions on Asian and Low-Resource Language Information Processing*, Vol. 23, No. 1, pp. 1-20, 2024.
[22] Y. Liu, Y. N. Xie, W. G. Li, X. He, H. G. He, L. B. Chen, and Q. Shen, "A Machine Learning-Based Risk Prediction Model for Post-Traumatic Stress Disorder During the COVID-19 Pandemic", *Medicina*, Vol. 58, No. 12, p. 1704, 2022.
[23] G. Blanco and A. Lourenço, "Optimism and Pessimism Analysis Using Deep Learning on COVID-19 Related Twitter Conversations", *Information Processing & Management*, Vol. 59, No. 3, p. 102918, 2022.
[24] H. V. Nguyen and H. Byeon, "Explainable Deep-Learning-Based Depression Modeling of Elderly Community after COVID-19 Pandemic", *Mathematics*, Vol. 10, No. 23, p. 4408, 2022.
[25] H. Li, E. Zheng, Z. Zhong, C. Xu, N. Roma, S. Lamkin, ... and W. Xu, "Stress Prediction Using Micro-EMA and Machine Learning During COVID-19 Social Isolation", *Smart Health*, Vol. 23, p. 100242, 2022.
[26] M. H. Al Banna, T. Ghosh, M. J. A. Nahian, M. S. Kaiser, M. Mahmud, K. A. Taher, ... and K. Andersson, "A Hybrid Deep Learning Model to Predict the Impact of COVID-19 on Mental Health from Social Media Big Data", *IEEE Access*, Vol. 11, pp. 77009-77022, 2023.
[27] J. Wu, X. Wu, Y. Hua, S. Lin, Y. Zheng, and J. Yang, "Exploring Social Media for Early Detection of Depression in COVID-19 Patients", In: *Proc. of the ACM Web Conf.*, pp. 3968-3977, 2023.
[28] N. Liu and J. Zhao, "A BERT-Based Aspect-Level Sentiment Analysis Algorithm for Cross-Domain Text", *Computational Intelligence and Neuroscience*, Vol. 2022, pp. 1-11, 2022.
[29] X. Wen, X. Liu, C. Yu, H. Gao, J. Wang, Y. Liang, ... and Y. Bai, "IOOA: A Multi-Strategy Fusion Improved Osprey Optimization Algorithm for Global Optimization", *Electronic Research Archive*, Vol. 32, No. 3, pp. 2033-2074, 2024.