
Research article

Modified arithmetic optimization algorithm with Deep Learning based data analytics for depression detection

Nuha Alruwais¹, Hayam Alamro^{2,*}, Majdy M. Eltahir³, Ahmed S. Salama⁴, Mohammed Assiri⁵ and Noura Abdelaziz Ahmed⁵

¹ Department of Computer Science and Engineering, College of Applied Studies and Community Services, King Saud University, Saudi Arabia, P.O.Box 22459, Riyadh 11495, Saudi Arabia

² Department of Information Systems, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia

³ Department of Information Systems, College of Science & Art at Mahayil, King Khalid University, Saudi Arabia

⁴ Department of Electrical Engineering, Faculty of Engineering & Technology, Future University in Egypt, New Cairo 11845, Egypt

⁵ Department of Computer Science, College of Sciences and Humanities- Aflaj, Prince Sattam bin Abdulaziz University, Aflaj 16273, Saudi Arabia

* **Correspondence:** Email: hmamro@pnu.edu.sa.

Abstract: Depression detection is the procedure of recognizing the individuals exhibiting depression symptoms, which is a mental illness that is characterized by hopelessness, feelings of sadness, persistence and loss of interest in day-to-day activities. Depression detection in Social Networking Sites (SNS) is a challenging task due to the huge volume of data and its complicated variations. However, it is feasible to detect the depression of the individuals by examining the user-generated content utilizing Deep Learning (DL), Machine Learning (ML) and Natural Language Processing (NLP) approaches. These techniques demonstrate optimum outcomes in early and accurate detection of depression, which in turn can support in enhancing the treatment outcomes and avoid more complications related to depression. In order to provide more insights, both ML and DL approaches possibly offer unique features. These features support the evaluation of unique patterns that are hidden in online interactions and address them to expose the mental state amongst the SNS users. In the current study, we develop the Modified Arithmetic Optimization Algorithm with Deep Learning for Depression Detection in Twitter Data (MAOADL-DDTD) technique. The presented MAOADL-

DDTD technique focuses on identification and classification of the depression sentiments in Twitter data. In the presented MAOADL-DDTD technique, the noise in the tweets is pre-processed in different ways. In addition to this, the Glove word embedding technique is used to extract the features from the preprocessed data. For depression detection, the Sparse Autoencoder (SAE) model is applied. The MAOA is used for optimum hyperparameter tuning of the SAE approach so as to optimize the performance of the SAE model, which helps in accomplishing better detection performance. The MAOADL-DDTD algorithm is simulated using the benchmark database and experimentally validated. The experimental values of the MAOADL-DDTD methodology establish its promising performance over another recent state-of-the-art approaches.

Keywords: data analytics; Twitter data; machine learning; depression detection; Nature-inspired algorithm; Computational Intelligence; Social networking

Mathematics Subject Classification: 11Y40

1. Introduction

Computational Intelligence (CI) is a subfield of Artificial Intelligence (AI) that concentrates on the development of intelligent algorithms and models accomplished through learning from data, adapting to changing environments and solving complex problems. The field of CI encompasses various techniques such as Machine Learning (ML), Swarm Intelligence (SI), Neural Networks (NNs), Fuzzy Logic (FL) and Genetic Algorithms (GAs). Psychological analysis refers to the process of extracting the psychological data from text-based data. In general, the text data is utilized to get rid of the judgement forming, emotions and opinions. Having views or opinions toward certain products or any subject is common human psychology that defines what one thinks of the topic or products [1]. These days, with the penetration of internet technology and social media, the way of sharing one's own opinions and expressing different emotions have drastically changed [2].

The public use social media sites, blogs and product review and recommendation websites to share their thoughts on various topics including political parties, products and movies on significant matters. Well-known social media platforms namely Reddit, Facebook and Twitter are the most used platforms in terms of sharing reviews and opinions about a product or a topic. Organizations and business firms utilize people-based psychological feedback to increase a product's quality and value [3]. Depression is often described as "a state of mind that expresses mood disorders like loss of appetite, depression, anxiety, unhappiness, lack of concentration, being bored and so on". It may badly affect life while at extreme situations, it eventually leads to suicidal activities. Depression remains the major cause of various health problems and affects people worldwide [4]. As discussed earlier, all individuals are prone to encounter depression and many do not get treatment since they are unaware of their depression status. Likewise, family members and friends have absolutely no idea about an individual's depression status. Those individuals who are at risk of depression commonly express themselves through social networking sites [5]. Social networking is a kind of interaction that fails to depend on facial expressions and eye contact, but it is expressed by commenting on pictures or messages. Therefore, social media data may help in detecting depression among people depending on the negative opinions posted by them.

Social media data is generated in huge volumes on a daily basis and is challenging to streamline [6]. Sentiment Analysis (SA) or opinion mining is a text analysis method that analyzes the thoughts of

human beings concerning organizations and the structures present in such entities. In SA, a feature can be a thing discussed by the persons about policies, services, events, products, individuals or entities [7]. The integration of features and respective sentiment words can produce high-quality, precise and meaningful SA outcomes. To segregate the text into negative, neutral or positive categories, the ML methods have been implemented in sentiment classification tasks [8]. In general, testing and training datasets are utilized in ML approaches. The training dataset is used to learn about the files whereas the testing datasets are utilized for validating the accomplishments of the ML methods [9]. ML methods encompass semi-supervised, supervised and unsupervised approaches. Among these, the unsupervised approaches are used for expression datasets but possess less predictive ability than the supervised approaches. Supervised approaches require data that are scarce and on well-known associates for training [10]. Though the semi-supervised approaches are trained with some communication data, its prediction outcomes are more opaque than the supervised algorithms.

In the current study, we develop the Modified Arithmetic Optimization Algorithm with Deep Learning for Depression Detection in Twitter Data (MAOADL-DDTD) technique. In the presented MAOADL-DDTD technique, the noise in the tweets is preprocessed through different ways. In addition to this, the Glove word embedding technique is also used to extract the features from the preprocessed data. For depression detection, the Sparse Autoencoder (SAE) model is used. The MAOA is used for optimum hyperparameter tuning of the SAE approach in order to increase the performance of the SAE model, which in turn helps in accomplishing enhanced detection performance. The novel part of the proposed framework is the MAOA-based parameter tuning that potentially resulted in improved accuracy and effectiveness in identifying the depression patterns in Twitter data. The MAOADL-DDTD method was simulated using the benchmark dataset and the outcomes established the superior performance of the proposed method.

2. Related works

Mbarek et al. [11] addressed the suicide prevention issue by identifying the suicidal profiles on social networks, especially Twitter. First, the author analyzed the twitter profiles and extracted different features from the data including account features relevant to the profile and parts relevant to the tweets. Second, the author presented the technique based on ML methods for detecting the suicidal profiles using the twitter data. Then, the author validated the proposed model using a profile dataset comprising persons who committed suicide earlier. In a study conducted earlier [12], a depression analysis and suicidal ideation recognition mechanism was presented to forecast the suicidal activities that reinforce the extent of depression. The researchers utilized the ML methods to detect the suicidal thoughts of a depressed twitter user from their tweets. They tested and trained the classifiers for distinguishing whether a user is depressed or not by utilizing feature extraction in its actions within the tweets. Priya et al. [13] predicted stress, anxiety and depression levels of a person utilizing ML methods i.e., five dissimilar Machine Learning algorithms. After that, the classes are imbalanced in the confusion matrix.

Alaskar and Ykhlef [14] framed a method that could categorize the tweets based on the depression features selected by the healthcare professionals. The author applied supervised ML methods for extracting the tweets with the most depression features. Then, for the identification of optimal settings for the proposed method, the author assessed the accuracy among the implemented supervised ML approaches. In literature [15], the authors aimed at detecting the depressed user posts using the ML methods. In this study, the NLP classified the mistreatment bidirectional encoder representations from the transformers BERT formula in order to observe the depression in an efficient and convenient

manner. Tadesse et al. [16] used NLP approaches and ML methods to train the data and evaluate the effectiveness of the presented technique.

Safa et al. [17] presented a new multimodal structure for the prediction of depression indicators from user profiles and introduced an automatic method to assess and gather the tweets depending on the self-reported statement. The author utilized automated image tagging, n-gram language models, bag-of-visual words and LIWC dictionaries. In the study conducted earlier [18], a new structure was proposed for efficiently finding the posts relevant to anxiety-and-depression, while preserving the semantic and contextual words utilized in the entire corpus at the time of implementing the BERT. Vayadande et al. [19] modeled an ML technique utilizing distant supervision so as to detect depression on twitter. The training data verified the twitter posts, involved with emojis that are categorized as noisy labels on a database. This study used multiple methods such as XGBoost, SVM, NB, LR and RF to differentiate the tweets as non-depressive and depressive.

Amanat et al. [20] presented a productive system by executing the LSTM approach, comprising two hidden states and a huge bias with Recurrent Neural Network (RNN) with two dense layers. This system was proposed to predict the depression from text as it is useful in protecting the individuals from mental disorders and suicidal affairs. Sardari et al. [21] introduced a structure utilizing endwise CNN-based Auto-encoder (CNN-AE) approach for learning extremely-relevant and discriminative features in raw sequential audio data. The aim was to identify the depressed people with high accuracy. Zogan et al. [22] examined explainable Multi-Aspect Depression Detection with Hierarchical Attention Network (MDHAN) for automatic detection of the depressed users on social network and explain the predictive model.

3. The proposed model

In the current study, a novel MAOADL-DDTD methodology has been developed for automated depression detection in twitter data. In the proposed MAOADL-DDTD technique, four subprocesses are involved namely, data preprocessing, Glove-based word embedding, SAE-based depression detection and MAOA-based parameter tuning. Figure 1 portrays the workflow of the MAOADL-DDTD methodology.

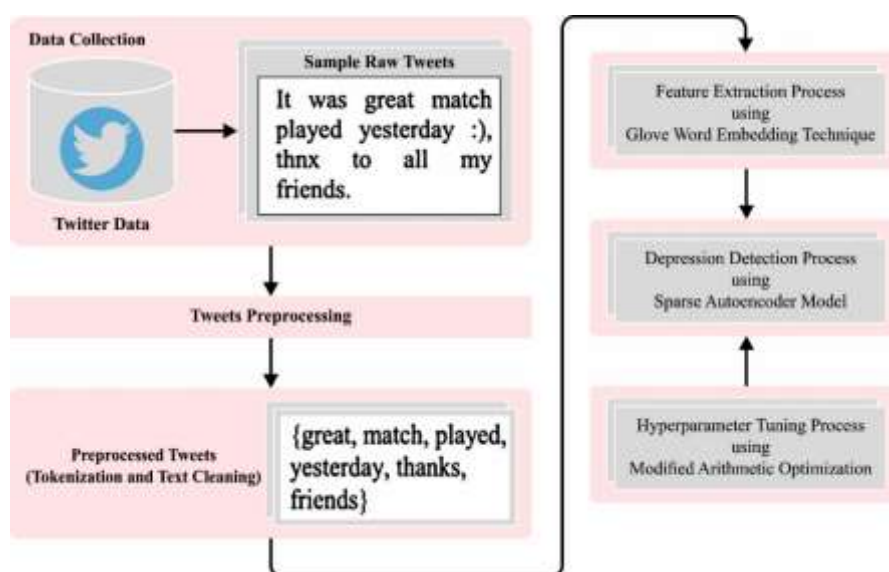


Figure 1. Workflow of the MAOADL-DDTD algorithm.

3.1. Data preprocessing

Text pre-processing is performed to clean the text information in such a way that the data is ready to be modeled for the next stage or process [23]. The steps of the pre-processing technique are briefed below:

Tokenization: In this phase, the phrases, symbols, words and other main entities (tokens) are separated from the text for additional investigation. This tokenization process breaks down the series of characters in a sentence (text) into units of words. Furthermore, the words or features are carefully chosen so that it is not valid. In such cases, punctuation marks and any other entities that are not letters are removed.

Transform Case: All the letters in the words are changed from uppercase to lowercase.

Filter Stopword: A significant word is taken from the token result. Now, both wordlist (saving significant words) and stoplist (discarding less significant words) algorithms are used.

Stopword Removal: This phase eliminates less significant words, including conjunction or articles, such as 'that', 'the', 'is', 'are', etc.

Generate N-grams: It is a fusion of adjectives that frequently appears to specify a sentiment of textual data with only one word. Trigrams have three words whereas Bigrams contain two words.

Stemming: This process is required to minimize the document's terms. Further, it is also used to group others words with similar meaning and a base word with different forms since they get various affixes.

3.2. Word embedding

In this work, the Glove word embedding technique is used. The relation of a word in the text, the context of the word in the text and words with similar semantics are captured by word embedding process [24]. In the word embedding technique, the words with similar meanings have the same representation. Every word is signified as a real-valued vector in the predetermined vector space. Instead of considering the local meaning, the GloVe word embedding technique considers the global context of the words. It is a pre-trained word embedding model that integrates the context window and global matrix factorization local models. A co-occurrence matrix X is created at this stage using a vocabulary of 400,000 words that are prevalently used. During the construction of X database, a context window is initialized to differentiate the left context from the right one, and the weighting function is reduced for the context window in such a way that the word pairs are d words apart from contributing $1/d$ to the overall count. Thus, the much-distant word pairs contribute less to the applicable data about word relationships.

The cost function is defined by Eq (1),

$$I = \sum_{i=1}^V \sum_{j=1}^V f(X_{ij})(\tilde{\omega}_i^T \tilde{\omega}_j + b_j + \tilde{b}_j - \log x_{ij})^2 \quad (1)$$

In Eq (2), \bar{w} and \tilde{w} show the vector representation of w_i and w_j correspondingly, w_j shows the set of context words of w_i , V denotes the vocabulary size and X_{ij} indicates the co-occurrence matrix. w_i shows a particular word in the message. b_i, \tilde{b} are biases and $f(X_{ij})$ denotes a weighting function that must adhere to the following properties.

$f(0) = 0$. If f is regarded as a continuous function, it must vanish as $x \rightarrow 0$ fast enough that the

$$\lim_{x \rightarrow 0} f(x) \log^2 x$$

is finite.

- $f(x)$ must be a non-decreasing value such that a rare co-occurrence is not over-weighted.
- $f(x)$ must be comparatively smaller for a larger value of x and so, the frequent co-occurrence is not over-weighted.

$$f(x) = \begin{cases} \left(\frac{x}{x_{\max}}\right)^\alpha & \text{if } x < x_{\max} \\ 1 & \text{otherwise} \end{cases} \quad (2)$$

The model performance weakly depends on the cut-off and is set as $x_{\max} = 100$ for each experiment.

3.3. Depression detection using SAE Model

For the depression detection process, the SAE approach is utilized in this study. AE is an unsupervised NN method that contains encoding and decoding layers; it is generally present in the architecture's input, output and Hidden Layers (HL), as represented in Figure 2 [25]. It maps the data to high-level parameter space for practical expression. Further, the feature data is recreated and improved. AE can moderate the size of the data and make sure the data features have invariance and integrity. This phenomenon ensures the quality of input feature instances to the succeeding networks that are trained by typical dimensionality reduction approaches like PCA and ICA.

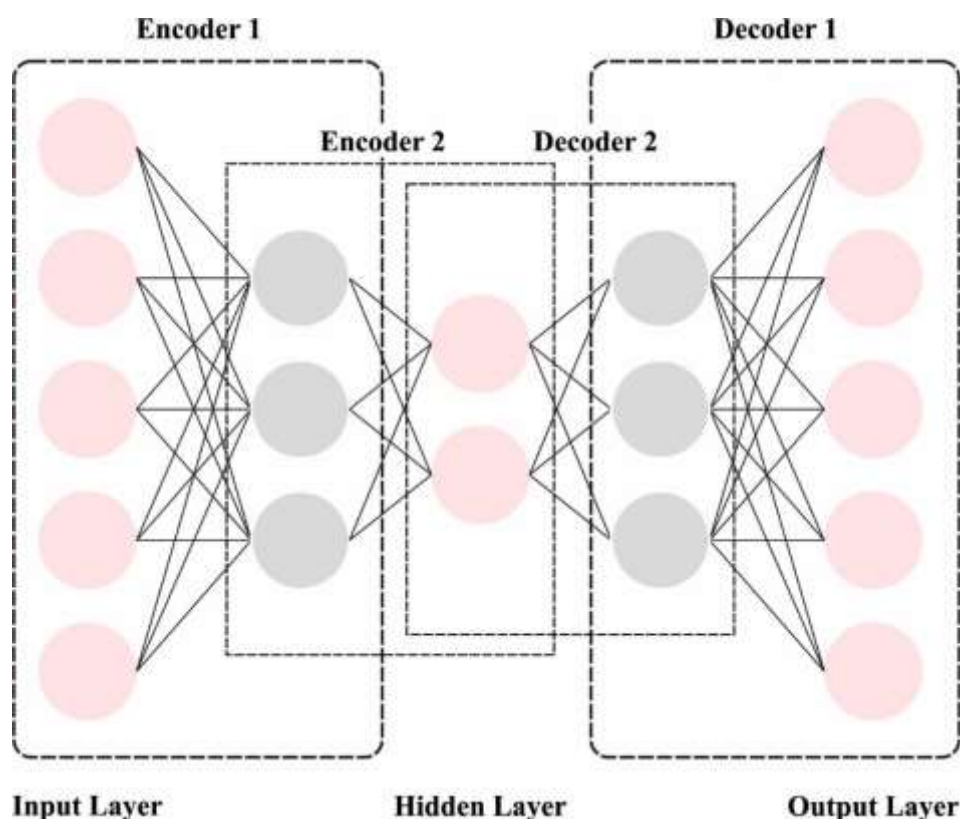


Figure 2. Architecture of the SAE.

SAE is highly reliable with AE with respect to structural composition. SAE improves a sparsity limit to the loss function of AE; moreover, only a few HL nodes can be ‘active’; therefore, the total AE network develops the sparse. Considering that the HL activation function utilizes sigmoid, if the outcome of HL is 0, then the node becomes ‘inactive’, and if the HL output is 1, then the node is ‘active’.

$$K_L(\rho||\hat{\rho}_j) = \rho \log \frac{\rho}{\hat{\rho}_j} + (1 - \rho) \log \frac{1-\rho}{1-\hat{\rho}_j} \quad (3)$$

$$\hat{\rho}_j = \frac{1}{m} \sum_{i=1}^m a_j(x_i) \quad (4)$$

Here, $\hat{\rho}_j$ implies the average sparse activation, x_j denotes the trained instance and m refers to the number of trained instances; $a_j(x)$ denotes the response outcome of the a j^{th} node of HL to the i^{th} sample. Generally, the sparsity co-efficient ρ is fixed at 0.05 or 0.1. Higher the KL divergence is, better the difference between ρ and $\hat{\rho}_j$ will be. If the KL divergence is equivalent to 0, then it denotes that both are totally equivalent.

$$JSAE(W, b) = J_{AE}(W, b) + \beta \sum_{j=1}^m K_L(\rho||\hat{\rho}_j) \quad (5)$$

Here, β denotes the weighted coefficient of the sparse constraints.

3.4. Hyperparameter Tuning using MAOA

In the final phase, the MAOA is exploited for optimal hyperparameter tuning of the SAE. AOA is a recent meta-inspired approach based on multiply (M), divide (D), add (A) and subtract (S) arithmetical operators [26]. Both add A and subtract S operations consider the dense data for the particle development and can prevent the algorithm from getting stuck in local optima. In the later iteration, it also prevents stagnation; the “multiply M and divide D operations” consider the dispersed data to identify the locations and particle exploration occurs closer to the optimum solution. Both exploration and exploitation processes in every iteration rely on the mathematically-optimized acceleration function (MOA).

$$MOA(iter) = MIN + iter \times \frac{MAX-MIN}{Max_iter} \quad (6)$$

In Eq (6), Max_iter corresponds to the iteration counts. MAX and MIN denote the maximal and minimal values of MOA , respectively; $iter$ shows the existing amount of iterations. The earlier formula is updated and formulated by Eq (7),

$$MOP(iter) = 1 - 1/Max_iter^{1/\delta} \quad (7)$$

In Eq (7), MOP denotes the mathematical optimization probability. δ represents the sensitive variable and is fixed as 5.

(1) Initialization

In the initialization stage, the lower boundary LB and the upper boundary UB of the particle position, the amount of particle dimensions, the number of particles and the maximal iterations are fixed. The location of the particle is randomly initialized, which assesses the optimal value and location

and also evaluates *MOA* and *MOP*.

(2) Exploration stage

When $r_1 > MOA$, then the exploration stage begins. r_1 corresponds to a randomly-generated value in the range of 0 and 1. The updated equation for the exploration process is given below.

$$x_{i,j}(iter + 1) = \begin{cases} best_j \div (MOP + \epsilon) \times w_j, r_2 < 0.5 \\ best_j \times MOP \times w_j, otherwise \end{cases} \quad (8)$$

In Eq (8), ϵ indicates a small positive number. $best_j$ shows the j -th dimension of the optimal particle location; r_2 represents the random integer in the range of 0 to 1.

(3) Exploitation stage

When $r_1 \leq MOA$, the algorithm enters the development stage. The location update equation is given herewith

$$x_{i,j}(iter + 1) = \begin{cases} best_j - (MOP + \epsilon) \times w_j, r_3 < 0.5 \\ best_j + MOP \times w_j, otherwise \end{cases} \quad (9)$$

In Eq (9), r_3 denotes the random value within [0, 1].

$$w_j = ((ub_j - lb_j) \times \mu + lb_j) \quad (10)$$

In Eq (10), μ is employed for fine-tuning the parameter of the search phase and is fixed at 0.5.

Though AOA produces highly competitive output, it has a few drawbacks like slower convergence and in the later stages, it tends to get stuck into the localization of the optimizer search. However, the AOA improves the convergence speed and optimization-seeking ability.

MOA is a crucial parameter that balances the local optimization task with global search. When the iteration count increases, *MOA* linearly increases from [0.2-1], thus making it hard to match the reality of the algorithmic optimizer. During the early iteration, *MOA* gets linearly increased from 0.2. This phenomenon makes the AOA of not being capable of exploring additional search space and accordingly, inadequate global exploration ability. *MOA* assumes a larger value in the later iteration, which results in slow convergence and limited local evolution. To resolve these problems, *MOA* is reformulated as follows.

$$MOA = (\text{Max} - \text{Min}) \times \left(\cos \left(0.5\pi \times \frac{Iter}{\text{Max_Iter}} \right) \right) + 0.5\text{Min} \quad (11)$$

To enhance the solution of the problems, the algorithm heavily relies on the optimum performance. In order to enhance the solution accuracy, the study presents the Differential Evolution (DE) technique in AOA to design the MAOA that produces new individual via mutation and crossover and enhance the search capability. The implementation phase is shown below.

$$\vec{X}_m(it) = \vec{X}_{c3}(it) + F \cdot [\vec{X}_{c1}(it) - \vec{X}_{c2}(it)] \quad (12)$$

In Eq (12), $\vec{X}_{c1}(it)$, $\vec{X}_{c2}(it)$ and $\vec{X}_{c3}(it)$ denote three randomly selected individuals; F represents the random value within [0.2, 0.8], and $\vec{X}_m(it)$ shows the individual after mutation.

Then, the present individual is crossed with mutated individuals to attain the novel individuals and is expressed as follows:

$$\vec{x}_c^l(it) = \begin{cases} \vec{x}_m^l(it)r_3 < P_c \\ \vec{x}_i^l(it)r_3 \geq P_c \end{cases} \quad (13)$$

In Eq (13), P_c denotes the crossover probability; r_3 represents the arbitrary in the range of 0 and 1. $\vec{x}(it)$ represents the l -th dimension of the existing individual; and $\vec{x}(it)$ denotes the l -th dimension of the variant individual.

$$\vec{X}_i(it) = \begin{cases} \vec{X}_c(it)f(\vec{X}_c(it)) < f(\vec{X}_i(it)) \\ \vec{X}_i(it)f(\vec{X}_i(it)) \leq f(\vec{X}_c(it)) \end{cases} \quad (14)$$

Here, $f(\vec{X}_c(it))$ and $f(\vec{X}_i(it))$ represent the fitness values before and after the mutation, correspondingly

The MAOA approach produces the Fitness Function (FF) to accomplish superior classification outcomes. It describes a positive integer to exemplify the good solution of the candidate results. During this case, the mitigation of the classifier's rate of errors is supposed to be FF, as shown in Eq (15).

$$\begin{aligned} \text{fitness}(x_i) &= \text{ClassifierErrorRate}(x_i) \\ &= \frac{\text{number of misclassified samples}}{\text{Total number of samples}} * 100 \end{aligned} \quad (15)$$

4. Performance validation

The proposed model was simulated using Python 3.8.5 tool on a PC configured with i5-8600k, GeForce 1050Ti 4GB, 16GB RAM, 250GB SSD and 1TB HDD. The parameter settings are given herewith; learning rate: 0.01, dropout: 0.5, batch size: 5, epoch count: 50 and activation: ReLU.

In this section, the depression detection performance of the MAOADL-DDTD system was validated using the depression tweets' database from Kaggle repository [27]. The dataset comprises 10,000 instances under two classes, as provided in Tables 1 and 2 shows a few sample tweets.

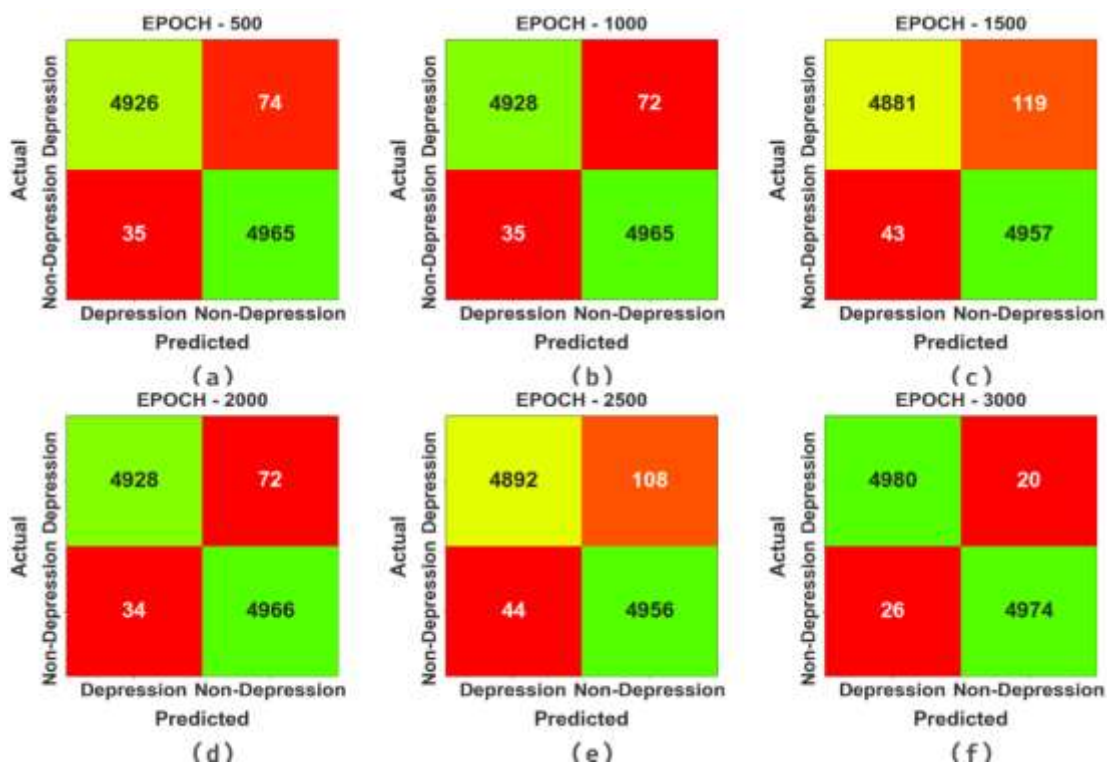
In Figure 3, the confusion matrices generated by the MAOADL-DDTD method under the varying number of epochs are shown. The figure shows that the MAOADL-DDTD method categorized the depression and non-depression sample tweets effectually.

Table 1 Details of database.

Class	No. of Instances
Depression	5000
Non-Depression	5000
Total Number of Instances	10000

Table 2. List of sample tweets.

Sample Tweets	Class Labels
damn taking this personality quiz and realizing have severe depression	1 (Depression)
damn louis did pull me out of my depression painting for the first time in months	1 (Depression)
my depression is kicking my ass right now so damn tired	1 (Depression)
feel like my night is going bad family calling me grumpy and whiny like so damn sorry for having depression	1 (Depression)
I'm trying to embrace it my depression needs to take a vacation for scorpio season damn	1 (Depression)
the reason we didn't work out years ago was so that we could have the chance today cool	0 (Non-Depression)
lol have Mars opposite Pluto and can do a lot of accents on some characters and sounds; it hidden talent have that practice at home lol ve always wanted to do voiceover acting	0 (Non-Depression)
same with all of them though it's just that wow so proud of them for never giving up and becoming the healing idols they dreamed to be	0 (Non-Depression)
Madoka had a bit of an attitude and is known for a bit of roasting like ena but they both mean well, and akari and minori are almost twins like they both started from the bottom and they made it to the top	0 (Non-Depression)
lol idk ever since daryl was smitten with connie he was looking a little cleaner in the last few episodes	0 (Non-Depression)

**Figure 3.** Confusion matrices of MAOADL-DDTD method (a-f) Epochs 500-3000.

In Table 3 and Figure 4, the overall depression classification outcomes of the MAOADL-DDTD methodology are provided. The outcomes show that the MAOADL-DDTD method distinguished the depression and non-depression tweet samples under several number of epochs. For instance, with 500 epochs, the MAOADL-DDTD technique gained an average $accu_y$ of 98.91%, $prec_n$ of 98.91%, $reca_l$ of 98.91%, F_{score} of 98.91% and an AUC_{score} of 98.91%. Meanwhile, with 500 epochs, the MAOADL-DDTD technique accomplished an average $accu_y$ of 98.91%, $prec_n$ of 98.91%, $reca_l$ of 98.91%, F_{score} of 98.91% and an AUC_{score} of 98.91%. Moreover, with 500 epochs, the MAOADL-DDTD method yielded an average $accu_y$ of 98.91%, $prec_n$ of 98.91%, $reca_l$ of 98.91%, F_{score} of 98.91% and an AUC_{score} of 98.91%. Lastly, with 500 epochs, the MAOADL-DDTD methodology produced an average $accu_y$ of 98.91%, $prec_n$ of 98.91%, $reca_l$ of 98.91%, F_{score} of 98.91% and an AUC_{score} of 98.91%.

Figure 5 shows the $accu_y$ outcomes of the MAOADL-DDTD algorithm in $train_g$ and $valid_n$ methods on the test database. The outcomes imply that the MAOADL-DDTD methodology attained enhanced $accu_y$ values over maximal number of epochs. Furthermore, the maximal $valid_n$ over $train_g$ $accu_y$ displays that the MAOADL-DDTD system attained its capability on the test database.

Table 3. Depression classification outcomes of the MAOADL-DDTD algorithm under varying number of epochs.

Class	$Accu_y$	$Prec_n$	$Reca_l$	F_{score}	AUC_{score}
Epoch - 500					
Depression	98.52	99.29	98.52	98.91	98.91
Non-Depression	99.30	98.53	99.30	98.91	98.91
Average	98.91	98.91	98.91	98.91	98.91
Epoch - 1000					
Depression	98.56	99.29	98.56	98.93	98.93
Non-Depression	99.30	98.57	99.30	98.93	98.93
Average	98.93	98.93	98.93	98.93	98.93
Epoch - 1500					
Depression	97.62	99.13	97.62	98.37	98.38
Non-Depression	99.14	97.66	99.14	98.39	98.38
Average	98.38	98.39	98.38	98.38	98.38
Epoch - 2000					
Depression	98.56	99.31	98.56	98.94	98.94
Non-Depression	99.32	98.57	99.32	98.94	98.94
Average	98.94	98.94	98.94	98.94	98.94
Epoch - 2500					
Depression	97.84	99.11	97.84	98.47	98.48
Non-Depression	99.12	97.87	99.12	98.49	98.48
Average	98.48	98.49	98.48	98.48	98.48
Epoch - 3000					
Depression	99.60	99.48	99.60	99.54	99.54
Non-Depression	99.48	99.60	99.48	99.54	99.54
Average	99.54	99.54	99.54	99.54	99.54

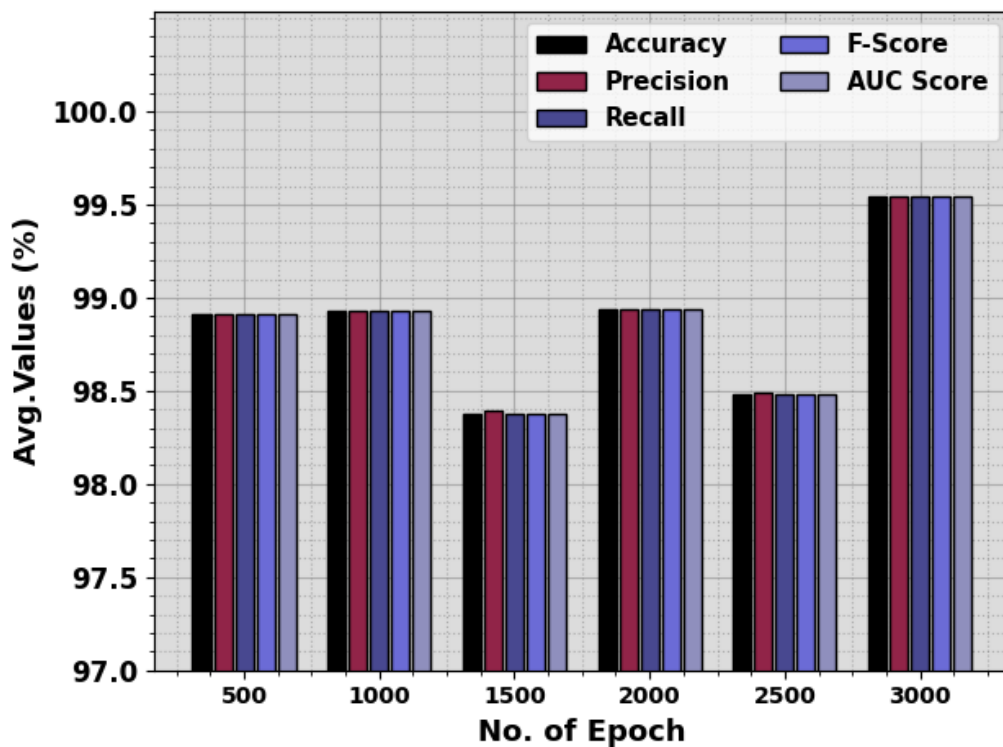


Figure 4. Average outcomes of the MAOADL-DDTD approach on varying number of epochs.

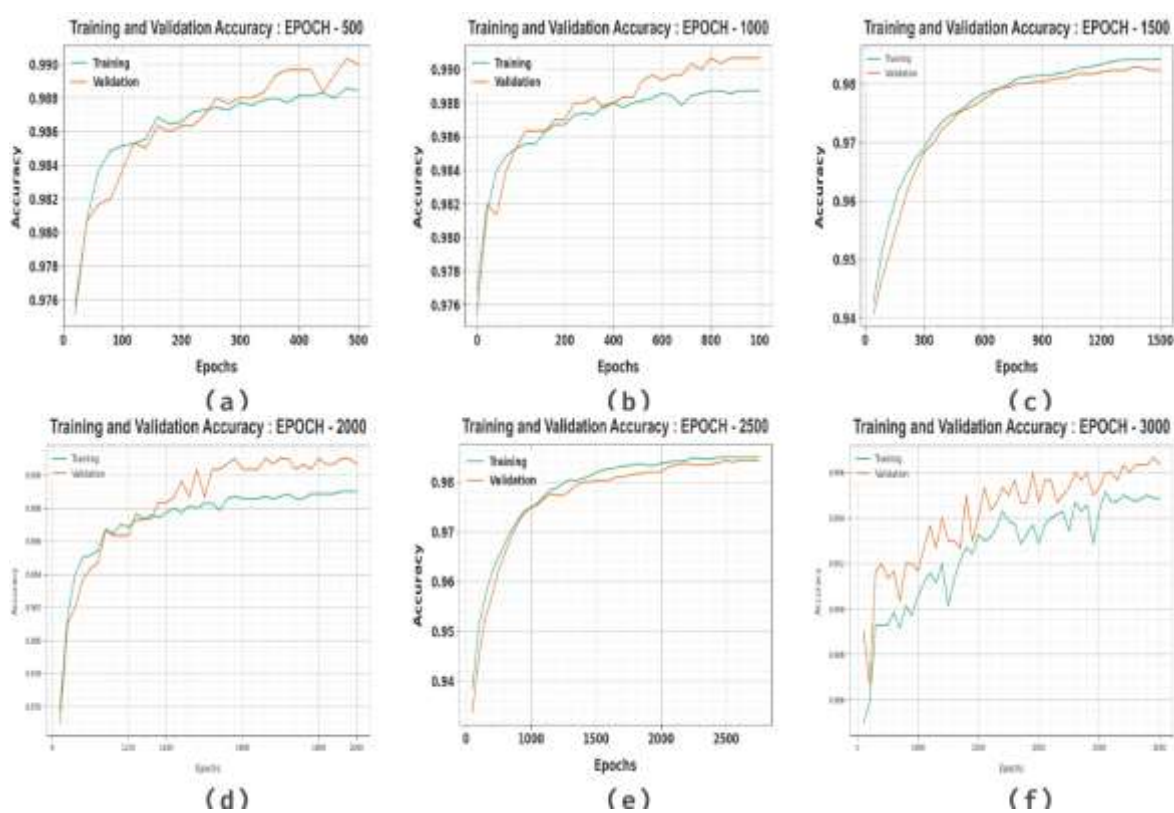


Figure 5. Accuracy curve of the MAOADL-DDTD methodology (a-f) Epochs 500–3000.

The loss outcomes of the MAOADL-DDTD algorithm at the time of $train_g$ and $valid_n$ on the test database are portrayed in Figure 6. The outcomes infer that the MAOADL-DDTD methodology obtained close $train_g$ and $valid_n$ loss values. It can be observed that the MAOADL-DDTD system obtained capability on the test database.

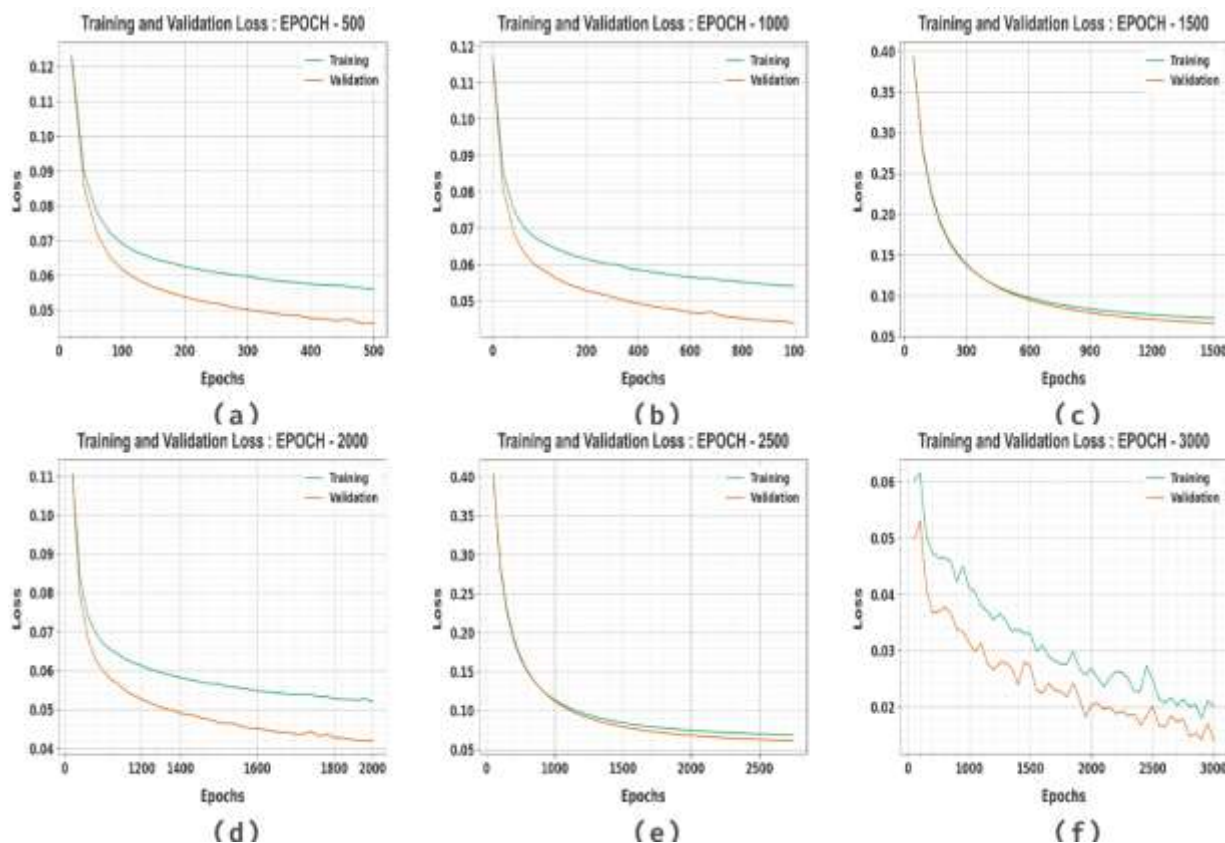


Figure 6. Loss curve of the MAOADL-DDTD methodology (a-f) Epochs 500-3000.

A comprehensive Precision-Recall (PR) analysis was conducted upon the MAOADL-DDTD methodology using the test database and the outcomes are revealed in Figure 7. The results portray that the MAOADL-DDTD system achieved enhanced PR values. Furthermore, it can be observed that the MAOADL-DDTD method obtained superior PR values on all the class labels.

In Figure 8, a ROC curve of the MAOADL-DDTD algorithm upon the test database is shown. The results define that the MAOADL-DDTD method achieved high ROC values. Additionally, the MAOADL-DDTD system obtained superior ROC values on all the class labels.

Finally, a widespread comparison analysis was conducted between the MAOADL-DDTD method with other approaches and the outcomes are provided in Table 4 and Figure 9 [28]. The outcomes indicate that the KNN, NB, CNN and MDL methods attained poor performance over other models. Furthermore, the SVM model gained certainly improved performance with an $accu_y$ of 97.21%, $prec_n$ of 72.40%, $reca_l$ of 63.20% and an F_{score} of 60.20%.

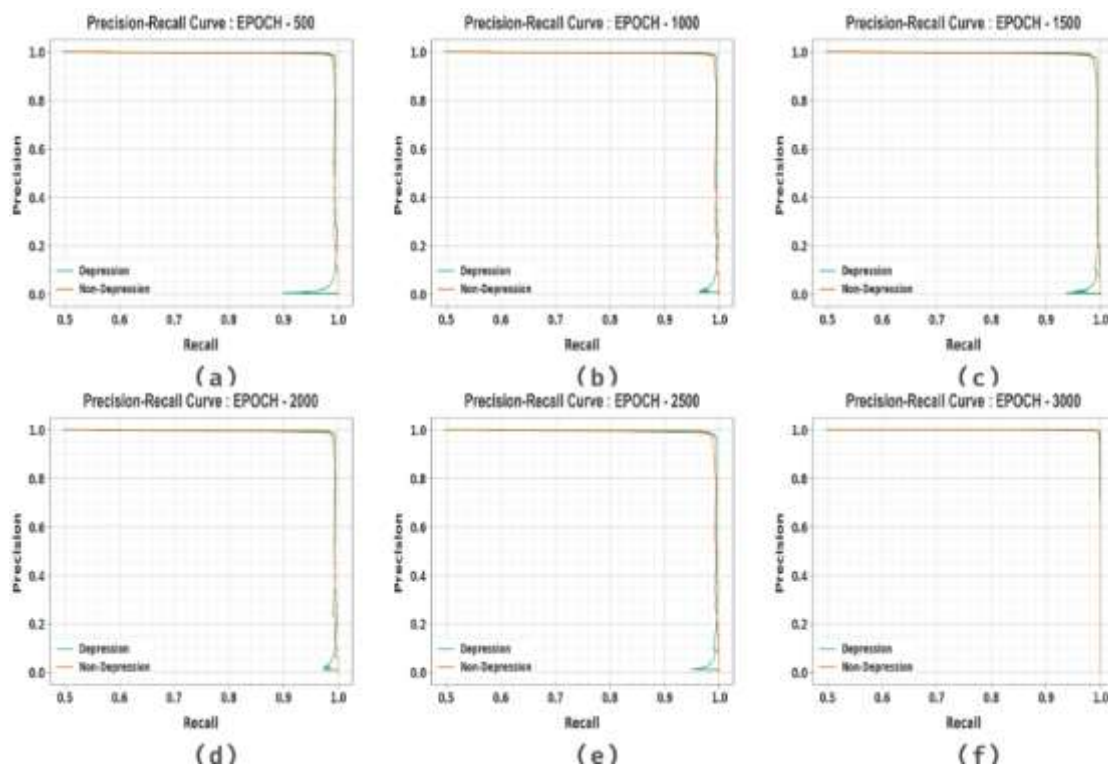


Figure 7. PR curve of the MAOADL-DDTD methodology (a–f) Epochs 500–3000.

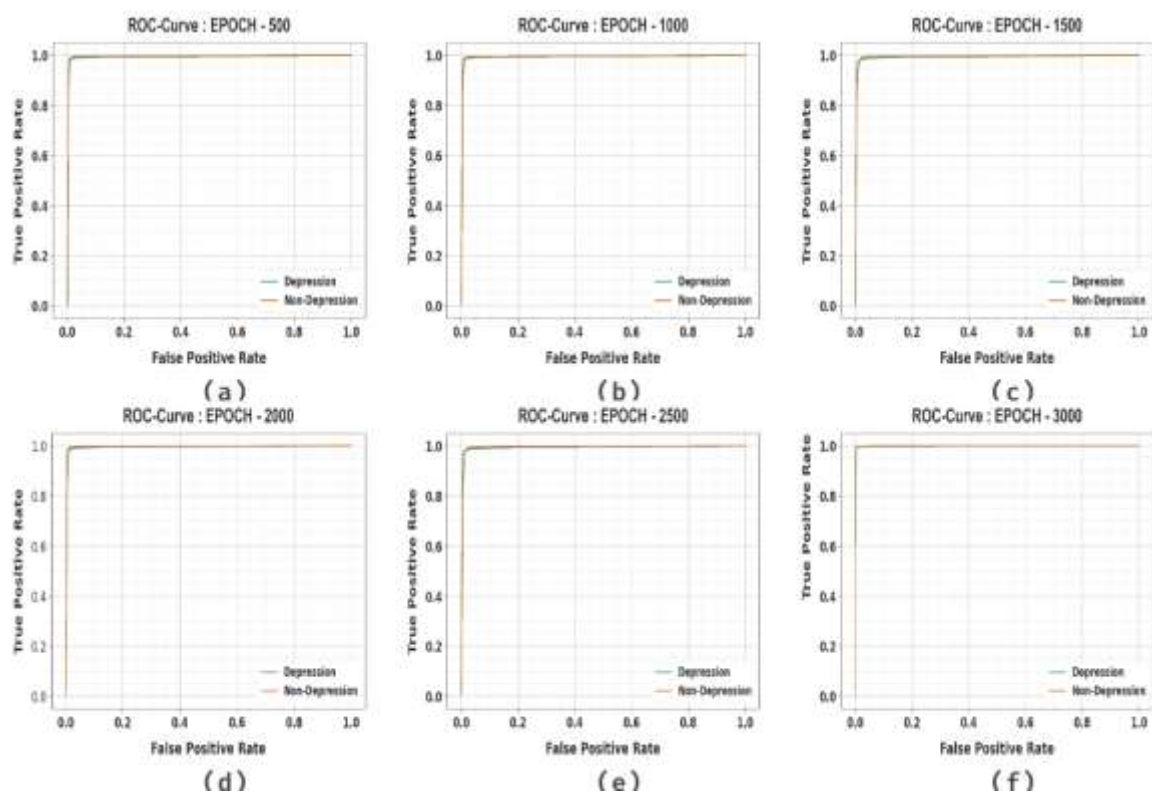


Figure 8. ROC curve of the MAOADL-DDTD method (a–f) Epochs 500–3000.

Table 4. Comparative analysis outcomes of the MAOADL-DDTD methodology and other recent algorithms [28].

Algorithms	$Accu_y$	$Prec_n$	$Reca_l$	$F1_{score}$
MAOADL-DDTD	99.54	99.54	99.54	99.54
KNN-Algorithm	82.10	81.60	94.00	93.90
NB-Algorithm	81.70	81.00	92.00	91.90
RF-Algorithm	98.60	98.50	99.00	99.00
CNN-Model	80.60	81.70	80.40	80.52
MDL-Model	78.70	79.00	78.62	78.98
SVM-Model	97.21	72.40	63.20	60.20

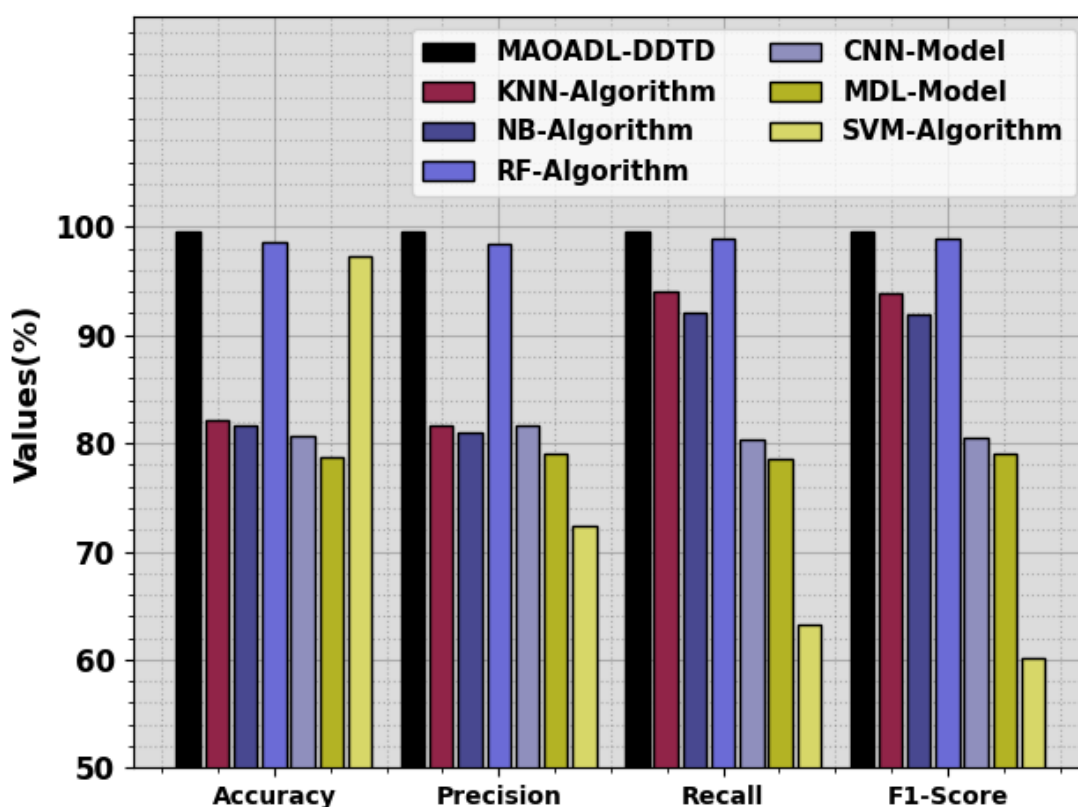


Figure 9. Comparative analysis outcomes of the MAOADL-DDTD method and other recent systems.

Along with that, the RF model resulted in further enhanced $accu_y$ of 98.60%, $prec_n$ of 98.50%, $reca_l$ of 99.00% and an F_{score} of 99%. However, the MAOADL-DDTD technique outperformed the rest of the models with the maximum $accu_y$ of 99.54%, $prec_n$ of 99.54%, $reca_l$ of 99.54% and an F_{score} of 99.54%. These outcomes confirmed the superior performance of the MAOADL-DDTD methodology over other existing methods under different measures.

5. Conclusions

In the current study, a novel MAOADL-DDTD methodology has been proposed for automated detection of depression from twitter data. In the proposed MAOADL-DDTD algorithm, four sub-processes are involved such as data preprocessing, Glove-based word embedding, SAE-based depression detection and MAOA-based parameter tuning. In the current research work, the MAOA is employed for optimum hyperparameter tuning of the SAE algorithm, which in turn helps in accomplishing better detection performance. The proposed MAOADL-DDTD algorithm was experimentally validated using the benchmark database. The experimental analysis results achieved by the MAOADL-DDTD method shows its promising performance over existing state-of-the-art approaches. Therefore, the MAOADL-DDTD technique has been found to be an effective tool for depression detection in a real-time environment. In the future, hybrid DL classification approaches can improve the outcomes of the MAOADL-DDTD method. A potential limitation of the proposed MAOADL-DDTD model is that it relies on publicly available twitter data, which may not fully represent the diverse and complex expressions of the scenario. Obtaining access to more comprehensive and private datasets could enhance the real-world applicability of the model. While the model incorporates MAOA for a hyperparameter tuning process, further research could explore advanced optimization techniques for fine-tuning the model using deep learning architecture and potentially improve the detection accuracy. In the future, extending the MAOADL-DDTD model to consider multi-modal data sources, such as images or user interactions, could provide a holistic understanding of depression in social media. This in turn enables highly comprehensive and accurate detection and classification.

Use of AI tools declaration

The authors declare that they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgment

The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University for funding this work through the large group Research Project under grant number (RGP2/29/44). Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2023R361), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. Research Supporting Project number (RSPD2023R608), King Saud University, Riyadh, Saudi Arabia. This study is supported by funding from Prince Sattam bin Abdulaziz University project number (PSAU/2023/R/1444). This study is partially funded by the Future University in Egypt (FUE).

Conflict of interest

The authors declare that they have no conflicts of interest. The manuscript was written with contributions from all authors. All authors have given approval to the final version of the manuscript.

References

1. S. Hinduja, M. Afrin, S. Mistry, A. Krishna, Machine learning-based proactive social-sensor service for mental health monitoring using twitter data, *Int. J. Inform. Manage. Data Insights*, **2** (2022), 100113. <https://doi.org/10.1016/j.jjime.2022.100113>
2. N. Semwal, M. Suri, D. Chaudhary, I. Gorton, B. Kumar, Multimodal Analysis and Modality Fusion for Detection of Depression from Twitter Data, *Association for the Advancement of Artificial Intelligence*, 2023, 1–5.
3. J. Stephen, P. Prabu, Detecting the magnitude of depression in Twitter users using sentiment analysis, *Int. J. Electr. Comput. Eng.*, **9** (2019), 3247. <https://doi.org/10.11591/ijece.v9i4.pp3247-3255>
4. A. Roy, K. Nikolitch, R. McGinn, S. Jinah, W. Klement, Z. A. Kaminsky, A machine learning approach predicts future risk to suicidal ideation from social media data, *NPJ Digit. Med.*, **3** (2020), 78. <https://doi.org/10.1038/s41746-020-0287-6>
5. T. Nijhawan, G. Attigeri, T. Ananthakrishna, Stress detection using natural language processing and machine learning over social interactions, *J. Big Data*, **9** (2022), 1–24. <https://doi.org/10.1186/s40537-022-00575-6>
6. G. Tiwari, G. Das, Machine learning based on approach for detection of depression using social media using sentiment analysis, *Depression*, **9** (2021), 10–16.
7. S. Ghosal, A. Jain, Depression and suicide risk detection on social media using fastText embedding and XGBoost classifier, *Procedia Comput. Sci.*, **218** (2023), 1631–1639. <https://doi.org/10.1016/j.procs.2023.01.141>
8. S. G. Burdisso, M. Errecalde, M. Montes-y-Gómez, A text classification framework for simple and effective early depression detection over social media streams, *Expert Syst. Appl.*, **133** (2019), 182–197. <https://doi.org/10.1016/j.eswa.2019.05.023>
9. P. Kumar, P. Samanta, S. Dutta, M. Chatterjee, D. Sarkar, Feature based depression detection from Twitter data using machine learning techniques, *J. Sci. Res.*, **66** (2022). <https://doi.org/10.37398/JSR.2022.660229>
10. R. Chiong, G. Budhi, S. Dhakal, F. Chiong, A textual-based featuring approach for depression detection using machine learning classifiers and social media texts, *Comput. Biol. Med.*, **135** (2021), 104499. <https://doi.org/10.1016/j.combiomed.2021.104499>
11. A. Mbarek, S. Jamoussi, A. Charfi, A. Hamadou, Suicidal Profiles Detection in Twitter. In: *WEBIST*, 2019, 289–296. <https://doi.org/10.5220/0008167602890296>
12. S. J. Pachouly, G. Raut, K. Bute, R. Tambe, S. Bhavsar, Depression detection on social media network (Twitter) using sentiment analysis, *Int. Res. J. Eng. Technol.*, **8** (2021), 1834–1839.
13. A. Priya, S. Garg, N. Tigga, Predicting anxiety, depression and stress in modern life using machine learning algorithms, *Procedia Comput. Sci.*, **167** (2020), 1258–1267. <https://doi.org/10.1016/j.procs.2020.03.442>
14. A. Alaskar, M. Ykhlef, Depression Detection from Arabic Tweets using machine learning techniques, *J. Comput. Sci. Soft. Devel.*, 2021, 1–10.
15. D. Solse, A. Magar, P. Harde, N. Palve, M. Jagatap, Depression detection by analyzing social media post in machine learning using bert algorithm, *Int. Res. J. Modernization Eng. Technol. Sci.*, **4** (2022).
16. M. Tadesse, H. Lin, B. Xu, L. Yang, Detection of depression-related posts in reddit social media forum, *IEEE Access*, **7** (2019), 44883–44893. <https://doi.org/10.1109/ACCESS.2019.2909180>

17. R. Safa, P. Bayat, L. Moghtader, Automatic detection of depression symptoms in twitter using multimodal analysis, *J. Supercomput.*, **78** (2022), 4709–4744. <https://doi.org/10.1007/s11227-021-04040-8>
18. K. Zeberga, M. Attique, B. Shah, F. Ali, Y. Z. Jembre, T. S. Chung, A novel text mining approach for mental health prediction using Bi-LSTM and BERT model, *Comput. Intell. Neurosci.*, 2022. <https://doi.org/10.1155/2022/7893775>
19. K. Vayadande, A. Bodhankar, A. Mahajan, D. Prasad, S. Mahajan, A. Pujari, et al., Classification of Depression on social media using Distant Supervision, In: *ITM Web of Conferences*, **50** (2022). EDP Sciences. <https://doi.org/10.1155/2022/7893775>
20. A. Amanat, M. Rizwan, A. R. Javed, M. Abdelhaq, R. Alsaqour, S. Pandya, et al., Deep learning for depression detection from textual data, *Electronics*, **11** (2022), 676. <https://doi.org/10.1155/2022/7893775>
21. S. Sardari, B. Nakisa, M. N. Rastgoo, P. Eklund, Audio based depression detection using Convolutional Autoencoder, *Expert Syst. Appl.*, **189** (2022), 116076. <https://doi.org/10.1016/j.eswa.2021.116076>
22. H. Zogan, I. Razzak, X. Wang, S. Jameel, G. Xu, Explainable depression detection with multi-aspect features using a hybrid deep learning model on social media, *World Wide Web*, **25** (2022), 281–304. <https://doi.org/10.1007/s11280-021-00992-2>
23. D. A. Kristiyanti, I. S. Sitanggang, S. Nurdianti, Feature selection using new version of V-Shaped transfer function for Salp Swarm Algorithm in sentiment analysis, *Computation*, **11** (2023), 56. <https://doi.org/10.3390/computation11030056>
24. S. Giri, S. Das, S. B. Das, S. Banerjee, SMS spam classification–simple deep learning models with higher accuracy using BUNOW and GloVe word embedding, *J. Appl. Sci. Eng.*, **26** (2023), 1501–1511.
25. Y. Liu, J. Kang, L. Wen, Y. Bai, C. Guo, Health status assessment of diesel engine valve clearance based on BFA-BOA-VMD adaptive noise reduction and multi-channel information fusion, *Sensors*, **22** (2022), 8129. <https://doi.org/10.3390/s22218129>
26. Y. Zou, R. Wu, X. Tian, H. Li, Realizing the improvement of the reliability and efficiency of intelligent electricity inspection: IAQA-BP Algorithm for anomaly detection, *Energies*, **16** (2023), 3021. <https://doi.org/10.3390/en16073021>
27. <https://www.kaggle.com/datasets/samrats/depressiontweets?select=train2Data.csv>
28. A. Nadeem, M. Naveed, M. Islam, H. Afzal, T. Ahmad, K. Kim, Depression detection based on hybrid deep learning SSCL framework using Self-Attention mechanism: An application to social networking data, *Sensors*, **22** (2022), 9775. <https://doi.org/10.3390/s22249775>



AIMS Press

© 2023 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)