

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/366893106>

# Deep Learning for Depression Detection Using Twitter Data

Article in *Intelligent Automation & Soft Computing* · January 2023

DOI: 10.32604/iasc.2023.033360

CITATIONS

24

READS

1,849

5 authors, including:



**Doaa Khafaga**

Princess Nourah bint Abdulrahman University

149 PUBLICATIONS 3,753 CITATIONS

[SEE PROFILE](#)



**Mohamed Abouhawwash**

Mansoura University

266 PUBLICATIONS 4,201 CITATIONS

[SEE PROFILE](#)

## Deep Learning for Depression Detection Using Twitter Data

Doaa Sami Khafaga<sup>1</sup>, Maheshwari Auvdaippan<sup>2</sup>, K. Deepa<sup>3</sup>, Mohamed Abouhawwash<sup>4,5</sup> and Faten Khalid Karim<sup>1,\*</sup>

<sup>1</sup>Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh, 11671, Saudi Arabia

<sup>2</sup>Department of Computational Intelligence, SRM Institute of Science and Technology, Kattankulathur, Chennai, 603203, India

<sup>3</sup>Department of Computer Science and Engineering, M. Kumarasamy College of Engineering, Karur, 639113, India

<sup>4</sup>Department of Mathematics, Faculty of Science, Mansoura University, Mansoura, 35516, Egypt

<sup>5</sup>Department of Computational Mathematics, Science, and Engineering (CMSE), College of Engineering, Michigan State University, East Lansing, MI, 48824, USA

\*Corresponding Author: Faten Khalid Karim. Email: fkdialdin@pnu.edu.sa

Received: 15 June 2022; Accepted: 01 August 2022

**Abstract:** Today social media became a communication line among people to share their happiness, sadness, and anger with their end-users. It is necessary to know people's emotions are very important to identify depressed people from their messages. Early depression detection helps to save people's lives and other dangerous mental diseases. There are many intelligent algorithms for predicting depression with high accuracy, but they lack the definition of such cases. Several machine learning methods help to identify depressed people. But the accuracy of existing methods was not satisfactory. To overcome this issue, the deep learning method is used in the proposed method for depression detection. In this paper, a novel Deep Learning Multi-Aspect Depression Detection with Hierarchical Attention Network (MDHAN) is used for classifying the depression data. Initially, the Twitter data was preprocessed by tokenization, punctuation mark removal, stop word removal, stemming, and lemmatization. The Adaptive Particle and grey Wolf optimization methods are used for feature selection. The MDHAN classifies the Twitter data and predicts the depressed and non-depressed users. Finally, the proposed method is compared with existing methods such as Convolutional Neural Network (CNN), Support Vector Machine (SVM), Minimum Description Length (MDL), and MDHAN. The suggested MDH-PWO architecture gains 99.86% accuracy, more significant than frequency-based deep learning models, with a lower false-positive rate. The experimental result shows that the proposed method achieves better accuracy, precision, recall, and F1-measure. It also minimizes the execution time.

**Keywords:** Depression detection; twitter data; tweets; deep learning; swarm intelligence; multi-aspect depression detection; prediction



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## 1 Introduction

Sadness is a key to spoiling the human memory system. Memory patients often experience deterioration in intellectual capacity, including understanding and recall. According to World Health Organization (WHO), depression or more than 300 million people, according to (WHO). Sadness can hurt one's psychological well-being, family responsibilities, and academic facilities at work and contribute to damage mitigation. Among the most common risk factors for depression or incapacity is depression [1]. Prevention and treatment for the symptoms of depression can significantly enhance the odds of managing depression. It also reduces the harmful effects of depression on a people's well-being, wellness, and social aspects.

Depression is the most common diagnosable disorder, which is more prevalent than that of other mental disorders environments worldwide. Depression identification is generally challenging since it necessitates trained psychiatrists' precise and comprehensive psychological evaluations. In a preliminary phase [2], interviews, surveys, self-reports, and evidence by friends and family members. Furthermore, it is pretty usual for persons struggling with mental health issues to avoid going to hospitals to seek aid from physicians during the initial stages of their illness.

It is critical to distinguish between depressed and non-depressed people using online social media. The person's emotional state is described through social media activity, interaction, and postings [3]. Nevertheless, their emotional state will remain strong, contributing to a delayed diagnosis of major depression. Medical consultations and questionnaire methods are performed by institutions or organizations [4]. Wherein mental evaluation forms have been used to determine mental illness diagnoses are currently the most widespread methods utilized.

The main motivation behind this research is to identify the emotion of the people using their chats and messages on social media. Today every person has their friends and families by connecting social media. It's interesting to analyze the emotional state of people using their message conversations and posts. This helps to know the depressed state of people and helps them to recover. Depression detection helps to identify mental state and reduce the pressure through counseling.

The information-sharing traditionally happens between humans and doctors directly in the olden days whereas now the internet is used to share information. Many research shows that discussing the moods and behaviors of individuals on a status update can offer health-related data [5–8]. These resources can help you find education and information for activities, including diagnosis, treatment, and compensation claims [9,10]. Next, by recommendations on scaling their strategy to reach many population needs in much less period, social media is used.

Deep learning has increasingly been used for a variety of challenges, including share market forecasts [11,12], vehicle traffic and disaster hazards forecasts [13–15], and psychological disorder identification [16]. Furthermore, deep learning was used accurately for detecting social networks with considerably improved returns than typical machine learning techniques. Digital data and documents are increasingly being acknowledged as trusted sources of information for clinical decisions [17]. Machine learning may diagnose depression using emotions or feelings using machine learning and Natural Language Processing (NLP) approaches. Using different databases, deep understanding has already been employed to investigate individual physical and psychological disorders [18,19]. Developing and executing irrational AI choices and a lack of justifications for the systems' behavior are the main concerns in this domain.

Furthermore, diagnosing symptoms of depression from quick texts is difficult. To support students with self-anticipated depression symptoms, we wish to develop an approach that can mechanically recognize depressed indicators from messages by using a text-based sampling. The main contribution of the proposed method is given below:

- We go into depression in-depth, including its causes, signs, and forms. This research focuses on textual documents processing and recognizing symptoms of depression.
- Deep learning of social media's linguistic, contextual, spatial, and structural component information is used to build a sad diagnosis system.
- To our understanding, it's the first study to combine many aspects of thematic, spatial, and semantically characteristics with word representations in deep learning to identify sadness.
- Comprehensive tests are carried out on the benchmarks depression tweeting dataset, demonstrating the efficiency of our suggested strategy over baseline models.
- The suggested MDH-PWO improves accuracy while simultaneously reducing execution time.

The rest of our research article is written as follows: Section 2 discusses the related work on various Depression Detection Methods and Deep learning methods. Section 3 shows the algorithm process and general working methodology of the proposed work. Section 4 evaluates the implementation and results of the proposed method. Section 5 concludes the work and discusses the result evaluation.

## 2 Related Work

Numerous researches on earlier diagnoses of sadness have employed artificial intelligence technologies such as machine learning. The author [20] proposed using N-gram language modeling and vectors with topic analysis to categorize the stress levels of created psychological features. The Continuous Bag of Words (CBOW) encoding technique has been proposed in [21] for detecting depression from a Twitter dataset. The use of a trained machine learning method to quantify predictive indicators for predicting post-traumatic depression or anxiety was explored in [22]. They used Twitter accounts as the primary research method to construct a design in every linguistic style. The author [23] suggested an approach for determining depression on social media using a deep neural network. For their research, they used the Twitter dataset.

In comparing the product range using convolutional neural networks. This relies on language information for mood forecasting. They were successful in executing cutting-edge work using the outlined technique. Most depressive study studies rely on text information or person-descriptive techniques that use posts on social media to identify attributes. Textual-based highlighting is used to highlight linguistic aspects in social media posts, including such words, parts of speech (POS), N-grams, and other language features. Topic descriptors, such as height, ethnicity, job status, wealth, drinking and drug use, tobacco, or client, are prioritized inside the descriptive-based highlighted method.

Algorithms that employ this form of research target specific linguistic aspects of the users' posting, which are retrieved as quantitative reasoning, such as count-based techniques. Those characteristics characterize the post's language features, as detailed. For example, the researchers offer classifiers to identify depression probability. The writer's specific purpose is to quantify the likelihood of user problems resulting from their posts on social media. For that purpose, the researchers gathered information on user pages on social media over a year before developing depression and extracted behavioral aspects. The examined related to social interaction, emotions, linguistic and language patterns, ego-system, and pharmacological intervention references in Tab. 1. The researchers acquire their information via Amazon Mechanical Turk utilizing crowd-sourcing jobs that are not sustainable techniques.

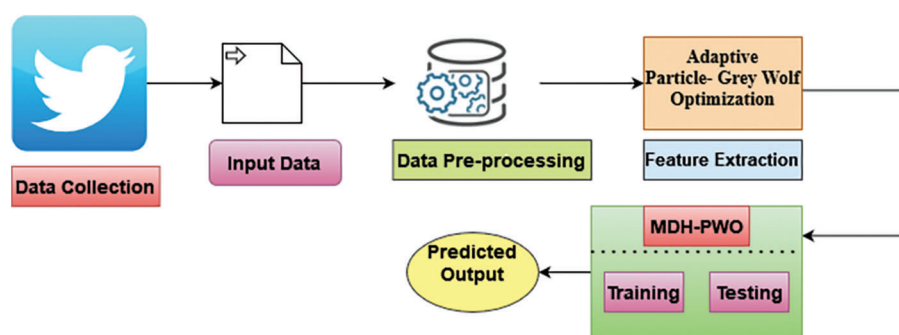
**Table 1:** Methodologies used for depression detection, discovery and constraints

Author	Methodologies used	Discovery	Constraints
U. Nisa et al. [24]	SVM, NB methods	NB provides excellent precision.	The information is in Arabic.
R. Chiong et al. [25]	MDL, MSNL, NB, WDL methods	MDL correctness is excellent.	Dedicated to user confessions
H. Fujita et al. [26]	DT, SVM, Ensemble methods	DT provides excellent precision.	The data focuses on the responses.
A. O. Ibitoye et al. [27]	KNN, SS3, NB, and SVM methods were used	Among other methods the greatest is SS3.	It is time-consuming
A. Trifan et al. [28]	SVM, DT, and NB were used	SVM has a high degree of precision.	Overfit information is unavoidable

The issues studied in the above literature works are discussed here. KNN and SVM are good ML algorithms for feature extraction, clustering, and classification. Depression detection needs some text-based processing in their projects. This results in overfitting, false predictions, and a time-consuming process. to overcome this issue deep learning-based optimized algorithms are used.

### 3 Proposed Methodology

A novel deep learning-based Multi-Aspect Depression Detection with Hierarchical Attention Network (MDHAN). The Adaptive particle swarm and grey Wolf optimization methods (MDH-PWO) are used for optimal outcomes. It detects the user's depression state with the help of social media. Initially, it collects data, and the preprocessing is done. Next, the features are extracted, and then the MDH-PWO is applied for the training and testing process. Finally, the results are predicted. Fig. 1 shows the architecture of the proposed method.

**Figure 1:** Architecture of proposed method

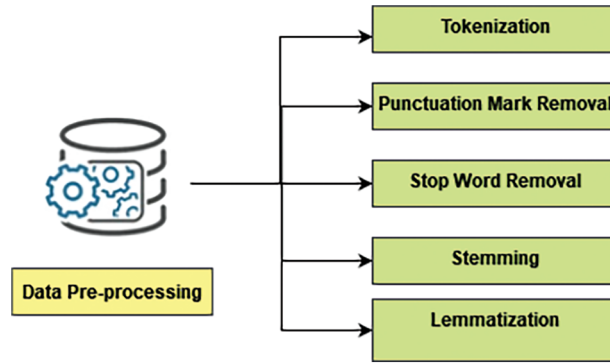
#### 3.1 Data Collection

Initially, the Twitter data is collected from Kaggle Website. It consists of a group of tweets [2].

#### 3.2 Pre-processing

In data pre-processing, it uses tokenization, punctuation mark removal, stops word removal, stemming, and lemmatization. Fig. 2 shows the following process involved in data pre-processing. URLs, tweets, and

stop-words were deleted from the database to clear it up. By separating the information into symbols or phrases, every row of the database was parsed. Following that, stemming and lemmatization was applied to the tokenized terms.



**Figure 2:** Steps involved in data pre-processing

### 3.3 Feature Extraction using Adaptive Particle-Grey Wolf Optimization

Particle Swarm Optimization (PSO) is a method designed by Eberhart and Kennedy that mimics the behavior of a flock of birds. It uses Adaptive Particle-Grey Wolf Optimization method. After extracting features, the extracted features are taken for feature selection. PSO, such as the evolutionary algorithm, finds an optimized response across stages. PSO, unlike Genetic Algorithms, does not include evolutionary modifications such as excellent hybridized mutations.

Similarly, Eberhart and Kennedy offer a theory regarding the herd's food hunting process during an environment in which each member of the group knows how much further away from the meal they had come and which location is nearest to the meal. The correct method to search for food is to follow the leader, who is closest to the feed in the flock. Researchers present a PSO method that can be used to modify this circumstance and resolve optimizing difficulties. The two factors that determine every constituent of the PSO are the object's present situation (XP) and velocities (VE).

Simultaneously moment, the fitness function calculates an optimal solution for every component. Just at the time of exit, the location of every element is stated as arbitrary. Every feature is affected by two items of data: the first is best, which would be the perfect location the component has ever had, and the other is guest, which would be the ideal location the entire flock has ever held. The PSO's details will navigate the issue area by adopting the best characteristics. After every step, the speed and direction of every element are determined using formulas (4) and (5), which are as described in the following:

$$VE_k^{t+1} = W * VE_k^t + C_1^t * rand * (pBest_k^t - X_k^t) + C_2^t * rand * (gBest - X_k^t) \quad (1)$$

$$X_k^{t+1} = X_k^t + VE_k^t \quad (2)$$

The numbers of C1 and C2 are usually specified as constant in PSO to equalize the research periods, more frequently to C1 = C2 = 1 or C1 = C2 = 2. In every cycle, an equation is used to change the accelerating constants. Eqs. (6) and (7) contain the additional co-efficient:

$$C_1^t = 1.2 - \frac{f(X_k^t)}{f(gBest)} \quad (3)$$

$$C_2^t = 0.5 + \frac{f(X_k^t)}{f(gBest)} \quad (4)$$

Here  $f(X_k^t)$  and  $C_1^t, C_2^t$  signify particles  $k$ 's efficiency and repetition  $t$ 's parameters, accordingly. Where  $f(gBest)$  is the swarm's global best efficiency. These values of 1.2 and 0.5 have been found through empirical work. The following is a description of the inertial equation:

$$W_t = (maxIter - t) * \frac{WMAX - WMIN}{maxIter} + WMIN \quad (5)$$

Sigmoid function is shown in Eq. (9)

$$V_{ij}^t(t) = sig(V_{ij}(t)) = \frac{1}{1 + e^{-V_{ij}^t(t)}} \quad (6)$$

A combination variation is a point mutation chance that leads to a modest number of GWO cycles in PSO's loop function. A point mutation frequency of 0.1 is applied in this scenario. This value is small enough the close circle is only reached just several times, with little impact just on swarm's optimal design stability.

### 3.4 Proposed (MDH-PWO) Multi-Aspect Depression Detection with Hierarchical Attention Network Using Deep Learning

Initially, it collects data from Twitter users. Then the user tweets and behaviors are pre-processed by tokenization, punctuation mark removal, stop word removal, stemming, and lemmatization. Next, the feature selection made by Adaptive Particle-Grey Wolf Optimization is discussed in Sections (3.2–3.3). Finally, the processed data are given into Multi-Aspect Depression Detection with Hierarchical Attention Network. The following sections explain the proposed work.

Assume that having a group GU of tagged consumers drawn from depression and non-depression data. Every tweet  $T_i$  has number of letters  $T_i = [W_{i1}, W_{i2}, \dots, W_{iN}]$ , whereby  $N$  is the maximum sum of words per message. Here MU be the overall amount of characteristics accessible to a user  $\{mu_i\}_{i=1}^{MU}$ , and let  $\{1, 2, \dots, S\}$  be the number of possible aspects characteristics, thus  $MU_s$  is the dimension of the  $S^{th}$  perspective. As a result, once we have a collection of linked user behaviors  $M$  and a set of user tweets  $AT$ , we can create a model. In Fig. 3 it shows the overall structure of proposed work. Our depression detection method looks like this:

$$f(AT, MU) \rightarrow \hat{z} \quad (7)$$

#### 3.4.1 RNN Encoder for User Tweets

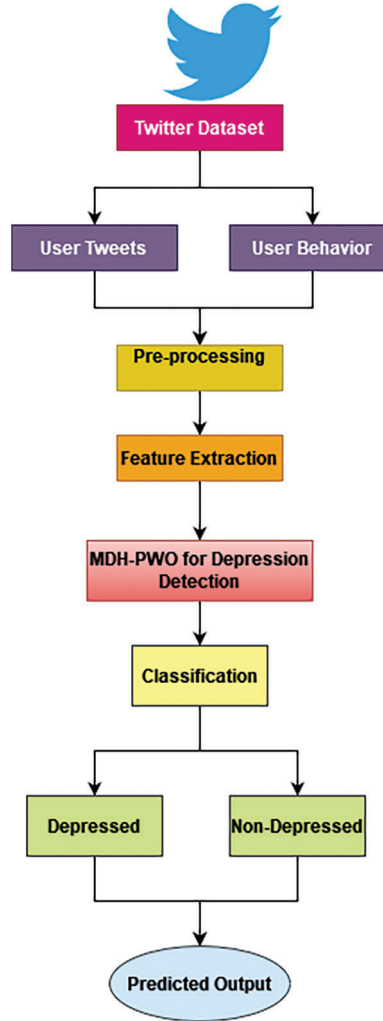
Studies have identified that HAN [2] may produce interpretations by evaluating the article's best significant words or phrases. A depressed person may post in various linguistic styles, including depressive language and references to pharmaceuticals and symptoms that can aid in the detection of depression. A social media post may also include linguistic cues at multiple phrases and tweet levels. To comprehend a depressed user on social media, each expression in a tweet and every person's post is significant.

#### 3.4.2 Encoder Used in Multi-Aspect

Consider the inputs that mimic user behavior as  $[mu_1, mu_2, \dots, mu_M]$ , wherein MU is the entire amount of characteristics and  $M_s$  is the dimensionality of the  $S^{th}$  aspect. As a result, to extract fine-grained data using user behavior characteristics, the multi-aspect characteristics are put throughout a one-layer MLP to acquire  $m_i$ :

$$q_i = nf\left(b + \sum_{i=1}^{MU} W_i mu_i\right) \quad (8)$$

here  $nf$  is the non-linear function as well as the result of behavioral modification is  $q_i$  is a high-level form which integrates behavioral information content and is useful in the diagnosis of sadness.



**Figure 3:** Overall structure of proposed work

### 3.4.3 Classifying the Data into Depressed or Non-Depressed

We have to estimate if the person is depressed or not in the classification model. Thus far, we've discussed how to encapsulate users' multi-aspect behavior traits ( $b$ ) or how to encapsulate users tweets by representing hierarchy organization at the phrase and tweeting layers ( $l$ ). The feature set of user behavior characteristics and users tweet is then constructed using both elements:

$$b = b_1, b_2, \dots, b_M \in R^{id \times M} \quad (9)$$

$$l = l_1, l_2, \dots, l_M \in R^{2d \times n} \quad (10)$$



We then combine those parts, which would be indicated as  $[b, l]$ . A sigmoid layer is often used to classify the outputs of this system:

$$\hat{y} = \text{sigmoid}(p_{nf} = [b, l]W_{nf}) \quad (11)$$

Here  $\hat{y}$  is the expected probability vectors and  $y_0$  and  $y_1$  represent the forecasted likelihood of the label becoming 0 (not depressed) and 1 (depressed user), accordingly. Furthermore, for every person with ground-truth labeling  $y$ , we want to minimize the cross-entropy error:

$$\text{Loss} = - \sum_i y_i \cdot \log \hat{y}_i \quad (12)$$

#### 4 Result Analysis

The evaluation of datasets is essential for verifying and assessing the quality of any detection technique. A good dataset is necessary for providing reliable and effective outcomes. We used a Tweets-Scraped dataset, which contains over 4000 tweets and is publicly available on Kaggle. There are both sad and non-depressed tweets in the sample. Here are the benefits of implementing a deep recurrent neural network method to the tweeter dataset. To verify the developed method, we used an Intel(R) Core-i7 CPU with a speed of 3.0 GHz, 64 GB of storage, Windows 10 OS, TensorFlow with Keras library (deep learning tools), and TensorFlow with Keras library (deep learning tools). We divide the entire dataset into 80 percent and 20% components for training examples and testing established procedures. To improve the efficiency of our prediction method, we evaluate the accuracy, recalls, support, precision, and F-1 measure. The parameters used for assessment are described below.

##### 4.1 Accuracy

It is used to evaluate the classification of depressed and non-depressed tweets accurately.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (13)$$

##### 4.2 Precision

$$\text{Precision} = \frac{TP}{TP + FP} \times 100 \quad (14)$$

##### 4.3 Recall

$$\text{Recall} = \frac{TP}{TP + FN} \quad (15)$$

##### 4.4 F-Measure

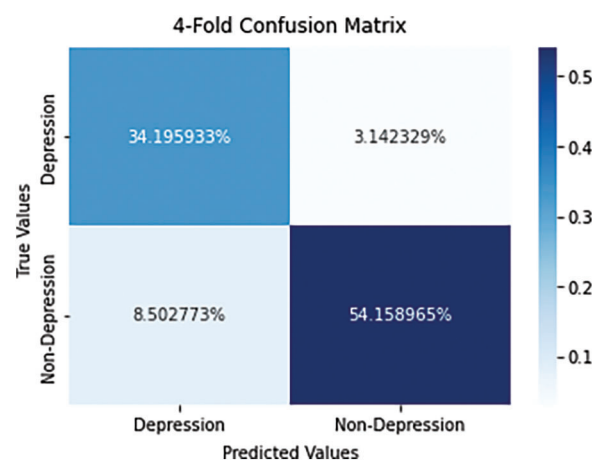
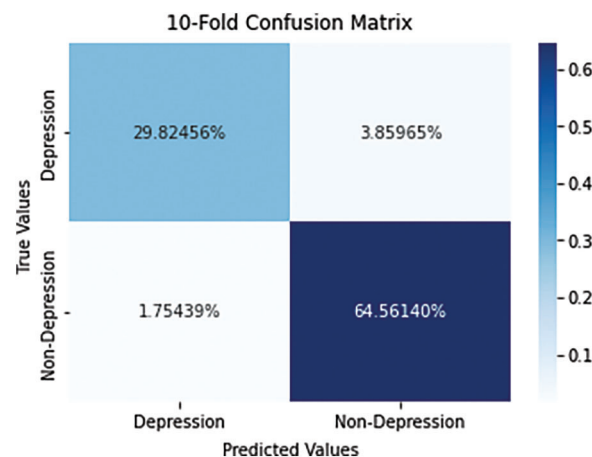
$$F \text{ Measure} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (16)$$

The dataset was subjected to four-fold and ten-fold cross-validation, with 10-fold cross-validation yielding the best accuracy of 99.86 percent. [Tab. 2](#) illustrates the results of the experiment of 10-fold cross-validation on the dataset that use the proposed methods for categorization. Depressed tweeting had a precision of 99.15 percent, while non-depressed tweets had 98.25 percent.

**Table 2:** Experimental results

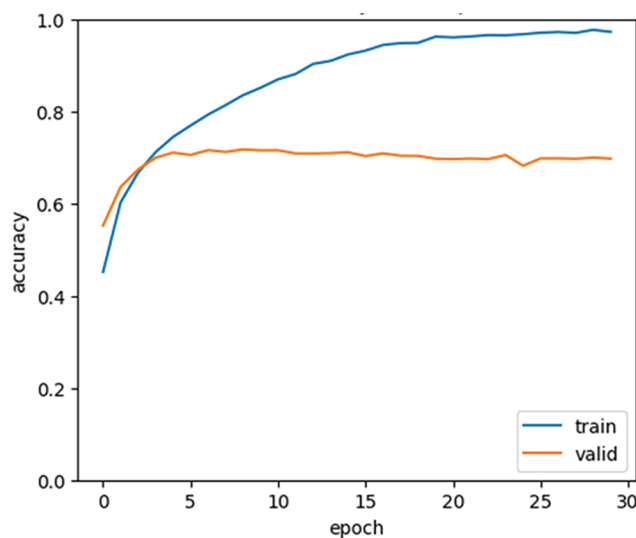
Metrics	Total number of tweets	Depression	Non-depression
Accuracy	4000	99.86%	97.65%
Precision	4000	99.15%	98.25%
Recall	4000	97%	98%
F1-measure	4000	99.32%	98.10%

Here 80% of the data was used for training, and 20% was used for testing in the testing phase. Figs. 4 and 5 show a 4-fold and 10-fold cross-validation confusion matrix [2] for the suggested approach.

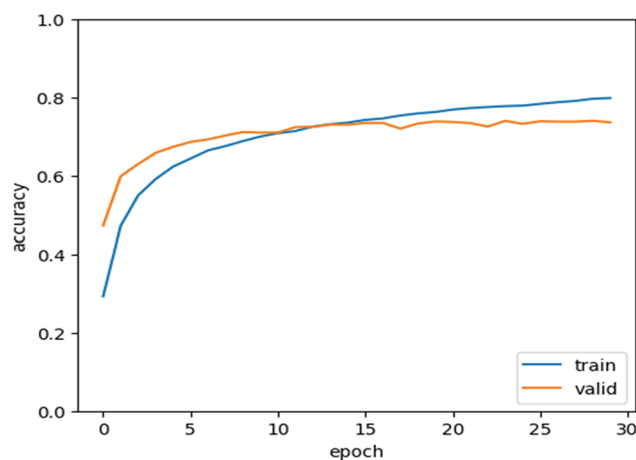
**Figure 4:** 4-Fold confusion matrix**Figure 5:** 10-Fold confusion matrix

The suggested methodology outperforms the tweets and produces the most accurate outcomes. The MDH-PWO maintains the sentences captured in storage forever, which allows it to analyze terms for detecting sadness in linguistic tweets. The trained model's overall performance is excellent and acceptable. The proposed framework's accuracy and epoch for the total number of instances are shown in

Figs. 6 and 7. The period indicates the training data cycle and the accuracy attained. Even as the epoch progresses, the accuracy improves. Figs. 6 and 7 show the 4-fold and 10-fold accuracy and age.



**Figure 6:** Accuracy and epoch using 4-fold



**Figure 7:** Accuracy and epoch using 10-fold

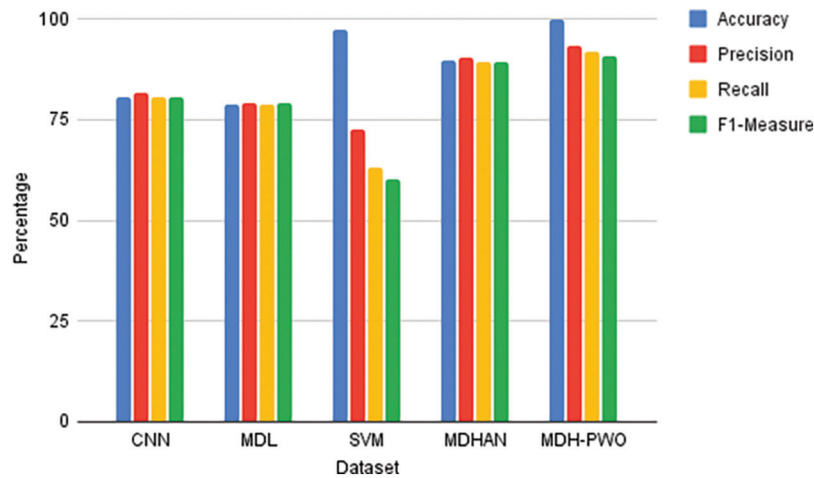
#### 4.5 Comparison and Discussions

The suggested method is compared with deep learning in textual data analysis. This research uses MDH-PWO technique on a text dataset of Norwegian young people's information on the internet channel. To detect depression, whereas we developed the system on tweets from tweeter dataset. Researchers extracted features using the linear discrimination method, but we applied PCA, an unsupervised learning methodology, to increase feature robustness and accuracy. Tab. 3 shows the various algorithms and parameters used in the method.

In Fig. 8 shows the comparisons of various algorithms used in detecting depression. The proposed method achieves better performance compared with other existing algorithms. It achieves 99.86% of accuracy, 93.45% of precision, 91.95% of recall, and 90.90% of F1 measure.

**Table 3:** Comparisons of various algorithms

Algorithms	Accuracy	Precision	Recall	F1-measure
CNN	80.6	81.7	80.4	80.52
MDL	78.7	79.0	78.62	78.98
SVM	97.21	72.4	63.20	60.2
MDHAN	89.5	90.2	89.35	89.3
MDH-PWO	99.86	93.45	91.95	90.90

**Figure 8:** Comparison of various algorithm

## 5 Conclusion

To manage depressed people in healthcare settings, it is necessary to recognize depression from text automatically. MDH-PWO was used to investigate multivariable human depression prediction strategies focused on a one-hot approach for robust features to describe depression and anxiety from textual information. A novel Deep Learning-based MDHAN with Adaptive Particle Swarm and grey Wolf optimization method is proposed. The Adaptive Particle Swarm and grey Wolf optimization methods are used for feature extraction. Finally, MDHAN detects the user submissions enhanced with Twitter's extra features. The characteristics used in this research will help machine learning make better decisions and contribute to a greater user experience for enhanced outcomes. Anxiety and depression recognition from the text could be used in behavioral healthcare centers for real-time model evaluation of common and severe depressive disorders. We plan to use the hybrid recurrent neural network on a massive dataset of mentally unstable patients in the future to study their behavior. The proposed method achieves better performance compared with other existing algorithms. It achieves 99.86% accuracy, 93.45% precision, 91.95% recall, and 90.90% F1 measured. In the future deep text analysis techniques can be used with swarm intelligence techniques.

**Acknowledgement:** Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2022R300), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

**Funding Statement:** This project is funded by Princess Nourah bint Abdulrahman University Researchers Supporting Project Number (PNURSP2022R300), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## References

- [1] A. Amanat, M. Rizwan, A. R. Javed, M. Abdelhaq, R. Alsaqour *et al.*, “Deep learning for depression detection from textual data,” *Electronics*, vol. 11, no. 5, pp. 676, 2022.
- [2] 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. 1–24, 2022.
- [3] T. M. Le, L. V. Tran and S. V. T. Dao, “A feature selection approach for fall detection using various machine learning classifiers,” *IEEE Access*, vol. 9, no. 4, pp. 115895–115908, 2021.
- [4] Y. Dong and X. Yang, “A hierarchical depression detection model based on vocal and emotional cues,” *Neurocomputing*, vol. 441, no. 1, pp. 279–290, 2021.
- [5] A. R. Javed, L. G. Fahad, A. A. Farhan, S. Abbas, G. Srivastava *et al.*, “Automated cognitive health assessment in smart homes using machine learning,” *Sustainable Cities and Society*, vol. 65, pp. 102572, 2021.
- [6] A. R. Javed, M. U. Sarwar, H. U. Khan, Y. D. A. Otaibi and W. S. Alnumay, “PP-SPA: Privacy preserved smartphone-based personal assistant to improve routine life functioning of cognitively impaired individuals,” *Neural Processing Letters*, vol. 13, no. 4, pp. 1–18, 2021.
- [7] M. Z. Uddin, K. K. Dysthe, A. Folstad and P. B. Brandtzaeg, “Deep learning for prediction of depressive symptoms in a large textual dataset,” *Neural Computing and Applications*, vol. 34, no. 1, pp. 721–744, 2022.
- [8] D. Who, *Other Common Mental Disorders: Global Health Estimates*, Geneva: World Health Organization, 2017.
- [9] A. E. Aladağ, S. Muderrisoglu, N. B. Akbas, O. Zahmacioglu and H. O. Bingol, “Detecting suicidal ideation on forums: Proof-of-concept study,” *Journal of Medical Internet Research*, vol. 20, no. 6, pp. e9840, 2018.
- [10] S. Wang, G. Peng, Z. Zheng and Z. Xu, “Capturing emotion distribution for multimedia emotion tagging,” *IEEE Transactions on Affective Computing*, vol. 12, no. 4, pp. 821–831, 2019.
- [11] J. M. Havigerová, J. Haviger, D. Kučera and P. Hoffmannová, “Text-based detection of the risk of depression,” *Frontiers in Psychology*, vol. 10, no. 5, pp. 513, 2019.
- [12] A. R. Javed, M. U. Sarwar, M. O. Beg, M. Asim, T. Baker *et al.*, “A collaborative healthcare framework for shared healthcare plan with ambient intelligence,” *Human-Centric Computing and Information Sciences*, vol. 10, no. 1, pp. 1–21, 2020.
- [13] W. Zehra, A. R. Javed, Z. Jalil, H. U. Khan and T. R. Gadekallu, “Cross corpus multi-lingual speech emotion recognition using ensemble learning,” *Complex & Intelligent Systems*, vol. 7, no. 4, pp. 1845–1854, 2021.
- [14] A. Ng and M. Jordan, “On discriminative vs. generative classifiers: A comparison of logistic regression and Naive Bayes,” *Advances in Neural Information Processing Systems*, vol. 14, no. 3, pp. 1–15, 2021.
- [15] H. Ni, S. Wang and P. Cheng, “A hybrid approach for stock trend prediction based on tweets embedding and historical prices,” *World Wide Web*, vol. 24, no. 3, pp. 849–868, 2021.
- [16] N. Liu, H. Yang and X. Hu, “Adversarial detection with model interpretation,” in *Proc. ACM SIGKDD Int. Conf. on Knowledge Discovery & Data Mining*, US, pp. 1803–1811, 2018.
- [17] N. Liu, M. Du and X. Hu, “Representation interpretation with spatial encoding and multimodal analytics,” in *Proc. ACM Int. Conf. on Web Search and Data Mining*, Melbourne, Australia, pp. 60–68, 2019.
- [18] H. Chen, Y. Li, X. Sun, G. Xu and H. Yin, “Temporal meta-path guided explainable recommendation,” in *Proc. ACM Int. Conf. on Web Search and Data Mining*, Texas, USA, pp. 1056–1064, 2021.
- [19] C. Y. Chiu, H. Y. Lane, J. L. Koh and A. L. Chen, “Multimodal depression detection on Instagram considering time interval of posts,” *Journal of Intelligent Information Systems*, vol. 56, no. 1, pp. 25–47, 2021.
- [20] S. Smys and J. S. Raj, “Analysis of deep learning techniques for early detection of depression on social media network-a comparative study,” *Journal of Trends in Computer Science and Smart Technology (TCSST)*, vol. 3, no. 1, pp. 24–39, 2021.

- [21] A. H. Orabi, P. Buddhitha, M. H. Orabi and D. Inkpen, “Deep learning for depression detection of Twitter users,” in *Proc. Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*, New Orleans, USA, pp. 88–97, 2018.
- [22] J. Kim, D. Lee and E. Park, “Machine learning for mental health in social media: Bibliometric study,” *Journal of Medical Internet Research*, vol. 23, no. 3, pp. e24870, 2021.
- [23] A. Wongkoblap, M. Vadillo and V. Curcin, “Depression detection of Twitter posters using deep learning with anaphora resolution: Algorithm development and validation,” *JMIR Mental Health*, vol. 14, no. 4, pp. 1–21, 2021.
- [24] U. Nisa and R. Muhammad, “Towards transfer learning using BERT for early detection of self-harm of social media users,” in *Proc. Conf. and Labs of the Evaluation Forum*, Bucharest, Romania, pp. 21–24, 2021.
- [25] R. Chiong, G. S. Budhi and S. Dhakal, “Combining sentiment lexicons and content-based features for depression detection,” *IEEE Intelligent Systems*, vol. 36, no. 6, pp. 99–105, 2021.
- [26] H. Fujita, A. Selamat, J. Lin and M. Ali, “Advances and trends in artificial intelligence: From theory to practice,” in *Proc. Int. Conf. on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, Kuala Lumpur, Malaysia, pp. 26–29, 2021.
- [27] A. O. Ibitoye, R. F. Famutimi, D. O. Olanloye and E. Akioyamen, “User centric social opinion and clinical behavioural model for depression detection,” *International Journal of Intelligent Information Systems*, vol. 10, no. 4, pp. 69, 2021.
- [28] A. Trifan, R. Antunes, S. Matos and J. L. Oliveira, “Understanding depression from psycholinguistic patterns in social media texts,” in *Proc. Conf. on Information Retrieval*, Stavanger, Norway, pp. 402–409, 2020.