International Journal of Engineering Trends and Technology
ISSN: 2231–5381 / https://doi.org/10.14445/22315381/IJETT-V73I6P108

Volume 73 Issue 6, 76-90, June 2025

Original Article

Dragonfly Optimised Regressive Gradient Convolutional Deep Belief Network for Depression Prediction using Social Media Texts

R. Geetha[1], D. Vimal Kumar[2]

[1,2]*Department of Computer Science, Rathinam College of Arts and Science, Coimbatore, Tamilnadu, India.*

[1]*Corresponding Author : gthram92@gmail.com*

Received: 19 March 2025 Revised: 12 May 2025 Accepted: 03 June 2025 Published: 28 June 2025

Abstract *- Depression is a widespread and deeply impactful mental health condition that affects millions of individuals globally. Conventional methods face challenges in improving the accuracy and robustness of depression detection. Dragonfly Optimized Piecewise Regression Gradient Convolutional Deep Belief Network (DOPR-GCDBN) model is proposed to enhance the accuracy of depression prediction through sentiment analysis of Twitter social media text data.*

Keywords *- Depression prediction, Twitter data, Convolutional deep belief network, Pre-processing, Dragonfly optimisation algorithm, Piecewise regression, Stochastic gradient.*

1. Introduction

A deep learning CNN-BiLSTM with an attention mechanism called CBA was developed in [1] to increase the accuracy of depression detection on social media text. However, the designed CBA model failed to focus on minimising the time complexity during the depression prediction. A Bidirectional Encoder Representation from Transformers with Convolutional Neural Network (BERT-CNN) Model was designed in [2] to provide the robust performance of depression detection across different social media contexts. However, it faced challenges in improving the accuracy and robustness of depression detection and optimising model performance and convergence. A semi-supervised learning approach was developed in [3] for detecting signs of depression from social media text. However, it did not achieve improved accuracy when handling larger datasets. A deep temporal clinical depression modelling approach was designed in [4] using Twitter posts by selecting clinically relevant features. However, it failed to analyse various model-feature combinations to improve the accuracy of depression detection further. An efficient, low-covariance multimodal combined with the spatio-temporal converter approach was developed in [5] to detect depression using acoustic and visual features from the social media data. However, the error minimisation was not efficiently handled during the depression detection. The Bidirectional Encoder Representations from Transformers (BERT) approach was designed in [6] to improve the performance of depression detection by extracting the Contextualized Embedding features. However, early detection of depression with minimal time was a major challenging issue. A Deep-Knowledge-aware model was introduced in [7] for detecting social media user's depression and their risk level. However, it failed to apply efficient techniques to extract information from text and other forms of digital trace data. A novel DeepFM network model was developed in [8] for depression detection in education students and minimising the loss function. However, the model failed to perform depression detection using social media texts. A multidimensional framework was designed in [9] to detect depression on social media by including textual, visual, behavioural, time series, and spatial features. However, the specificity of the framework was not improved. An ensemble of transformer-based models was designed in [10] to evaluate the severity of depression from social media posts into different classes. However, the model exhibited misclassifications when handling more data samples. An NLP-based system was developed in [11] to detect users' depression levels based on mental health information. However, it failed to identify the cases of mild and minimal depression. A Long Short-Term Memory (LSTM) was designed in [12] to detect depression on social media platforms through social media text data analysis. However, it failed to integrate multimodal data for early depression detection. A Bidirectional Encoder Representation from Transformers (BERT) model was developed in [13] to detect depression. However, the approaches were not applied to a broader range of datasets. Several machine-learning algorithms were designed [14] for early-stage depression detection using social media datasets. However, it failed to consider incorporating emotion recognition features to detect harmful content online and

academic stress. Machine learning techniques were developed in [15] based on structural and non-structural dual languages with the aim of depression detection. However, it failed to apply the advanced hybrid machine learning models to enhance the accuracy of depression prediction.

Table 1. Merits and demerits for various methods

S. No	Methods	Contribution	Merits	Demerits
1.	Deep learning CNN-BiLSTM with an attention mechanism called CBA	Deep learning CNN-BiLSTM with an attention mechanism called CBA was developed to increase the accuracy of depression detection on social media text	Accuracy was improved	The designed CBA model failed to minimise the time complexity during the depression prediction.
2.	Bidirectional Encoder Representations from Transformers with Convolutional Neural Network (BERT-CNN) Model	The BERT-CNN Model was used to provide the robust performance of depression detection across different social media contexts.	Depression detection time was reduced.	It Faced challenges in improving the accuracy and robustness of depression detection and optimising model performance and convergence.
3.	Semi-supervised learning approach	A semi-supervised learning approach was used for detecting signs of depression from social media text.	Precision was improved	It did not achieve improved accuracy when handling larger datasets
4.	Deep temporal clinical depression modelling approach	Deep temporal clinical depression modelling approach using Twitter posts by selecting clinically relevant features	Recall was improved	It failed to analyse various model-feature combinations to improve the accuracy of depression detection further.
5.	An efficient, low-covariance multimodal combined with the spatio-temporal converter approach	An efficient, low-covariance multimodal combined with the spatio-temporal converter approach was developed to detect depression using acoustic and visual features from the social media data.	Accuracy was improved	Error minimisation was not efficiently handled during the depression detection process.
6.	Bidirectional Encoder Representations from Transformers (BERT) approach	The Bidirectional Encoder Representations from Transformers (BERT) approach was used to improve the performance of depression detection by extracting the Contextualized Embedding features	Precision was improved	Early detection of depression with minimal time was a major challenging issue.
7.	Deep-Knowledge-aware model	The deep-knowledge-aware model was used for detecting the social media user's depression and their risk level.	Depression detection time was reduced.	Failed to apply efficient techniques to extract information from text, other forms of digital trace data
8.	DeepFM network model	The DeepFM network model was used for depression detection in education students and minimising the loss of function.	Accuracy was improved	The model failed to perform depression detection using social media texts.
9.	Multidimensional framework	A multidimensional framework was used to detect depression on social media by including textual, visual, behavioural, time series, and spatial features.	The error rate was minimised.	The specificity of the framework was not improved.
10.	Ensemble of transformer-based models	An ensemble of transformer-based models was used to evaluate the severity of depression from social media posts into different classes.	Precision was enhanced	The model exhibited misclassifications when handling a larger number of data samples.

1.1. Problem Statement

Depression Prediction has been given considerable attention. However, depression also leads to tiredness and poor concentration, even self-harm and suicide. Depression prediction to identify indicators of mental health issues in written text. The robust performance of depression prediction by using the CBA, however, does not address the time complexity involved in depression prediction. Bidirectional Encoder Representations from Transformers with Convolutional Neural Network (BERT-CNN) reduced the prediction time but failed to improve the accuracy. The method of BERT-CNN reduces the time consumption. However, it did not perform the stochastic gradient method to optimise the error rate further. To address the problem, the DOPR-GCDBN method is introduced, which enhances the accuracy of depression prediction.

1.2. Research Gap

Depression Prediction through sentiment analysis on social media is the method that helps to identify indicators of mental health issues in written or spoken text. A conventional deep learning CNN-BiLSTM with an attention mechanism called CBA for depression Prediction enhancement provides insufficient depression prediction accuracy. Some existing methods cannot focus on computation cost and classification error minimisation. To address these issues, a new technique called the DOPR-GCDBN model has been developed to enhance the accuracy of depression prediction through sentiment analysis of Twitter social media text data with minimum time consumption.

2. Related Works

An explainable multi-layer dynamic ensemble approach was designed in [16] to distinguish depression and evaluate its severity, aiming to improve diagnostic precision. However, the robust and interpretable framework was not applied to assess depression. A novel deep-learning transformers approach was developed in [17] to distinguish the various levels of depression using textual data. However, the error rate in the depression detection remained unaddressed. A Federated Learning (FL) model was developed in [18] to detect depression using social media posts. However, the scalability of FL was not improved in the learning process.

The NLP-based sentiment analysis method was developed in [19] to identify the high risk of depression on online social networks. However, it failed to implement an effective mechanism for enabling a deeper analysis of depression detection. A simple fuzzy inference model was introduced in [20] to predict depression levels based on sentiments and behaviors accurately. However, it failed to handle large sample-size datasets effectively. A new hybrid model was developed in [21] by integrating deep learning techniques with machine learning approaches for detecting depression from Twitter text data samples. However, it did not perform well on different social media networks. Integrating deep learning techniques and natural language processing models was developed in [22] to detect depression. However, it failed to consider other social media platforms like Twitter and Facebook. A new self-attention-based LSTM bidirectional TCN model was designed in [23] to differentiate suicidal ideation from social media posts.

However, it failed to experiment with different combinations of learning algorithms to achieve better performance. A transformer-based architecture was designed in [24] to distinguish and explain the appearance of depressive indications within user-generated content from social media. However, the hyperparameter optimisation remained unaddressed to enhance the model's accuracy. The XGBoost Classifier model was developed in [25] for depression detection through text embedding. However, it failed to analyse the various negative emotions and anxiety-based features for accurate depression detection.

A depression-related emotions classifier model was developed in [26] using Reddit user posts. Though the designed model increases the f1 score, an efficient and novel text classification approach was not employed to detect depression precisely. A novel machine learning models were implemented in [27] and used various embedding techniques to categorise stressful and non-stressful posts. However, the time complexity of classification was not minimised. A multi-label emotion graph representation model was developed in [28] for social media post-based mental health prediction.

However, the deep analysis model was not employed to enhance the accuracy. A Hierarchical Convolutional Neural Network (HCN) was developed in [29] for depression detection by extracting the fine-grained and relevant features on user historical posts. However, the error rate was not minimised. Attention-based models were developed in [30] for diagnosing depression indications from social media texts and labelled tweets. However, the complexity of depression detection was high. For finding the depression level of humans by machine, the model was developed in [31]. However, accuracy was not improved. A designed method was used to prevent suicide by identifying suicidal posts on social media was designed in [32] but failed to minimise the prediction time.

3. Proposal Methodology

Depression is a widespread mental health disorder that impacts individuals' lives. DOPR-GCDBN has been developed to improve the lives of individuals affected by depression.

3.1. Data Acquisition

Data acquisition involves gathering relevant and reliable data samples from the Depression: Twitter Dataset + Feature Extraction
https://www.kaggle.com/datasets/infamouscoder/mental-health-social-media

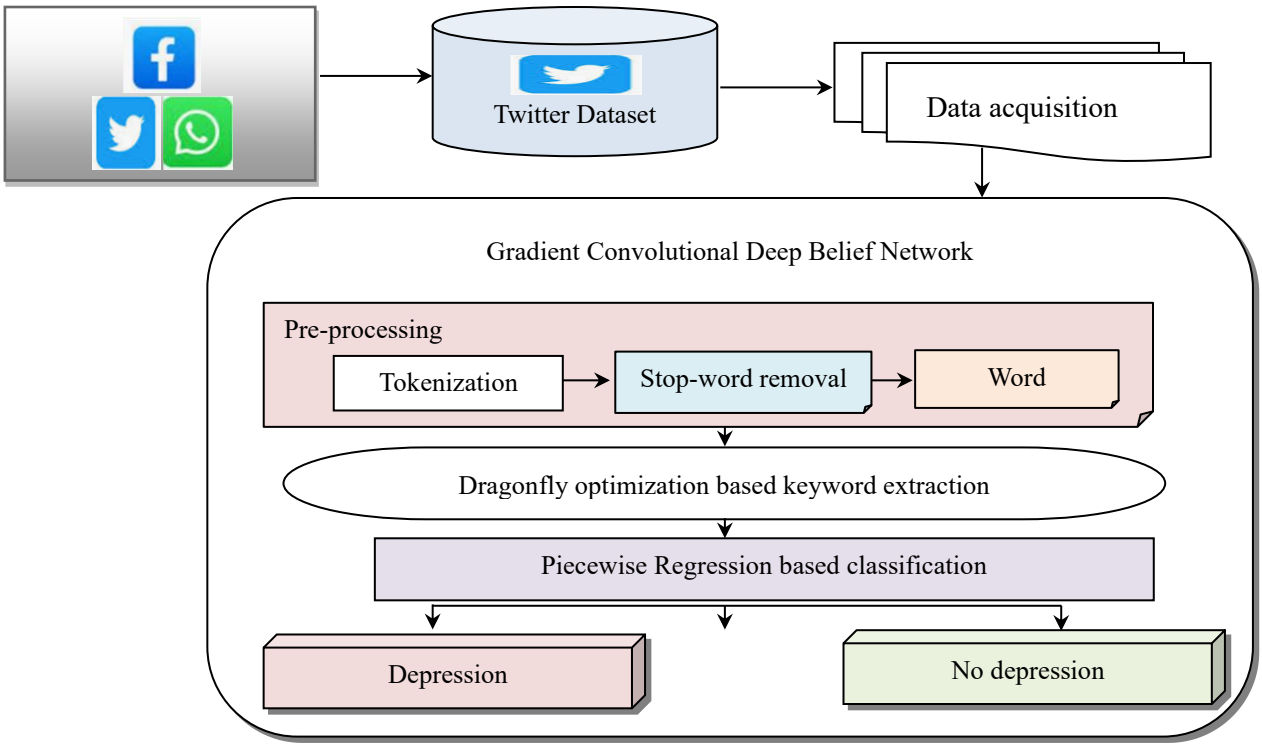

Fig. 1 Architecture diagram of DOPR-GCDBN model

Table 2. Attribute information

S. No	Features	Description
1	Index	
2	post_id	Post_id of the post
3	post_created	Post created on
4	post_text	Uncleaned Tweet
5	user_id	User Identification
6	followers	Number of Followers
7	friends	Number of Friends
8	favourites	Number of Favourites
9	statuses	Total Status Count
10	retweets	Total Retweets on the Current Tweet
11	label	Labels for Classification 1: depressed, 0:no depressed

3.2. Piecewise Regression-based Gradient Convolutional Deep Belief Network

CDBN is a deep artificial neural network comprising numerous layers of convolutional restricted Boltzmann machines combined to perform specific tasks.

The CDBN is a hierarchical generative model for integrating the different processes into their network structure and analysing the input samples, minimising dimensionality and complexity.

The DOPR-GCDBN model utilises CDBN to perform the depression prediction process for better accuracy and minimal time consumption.

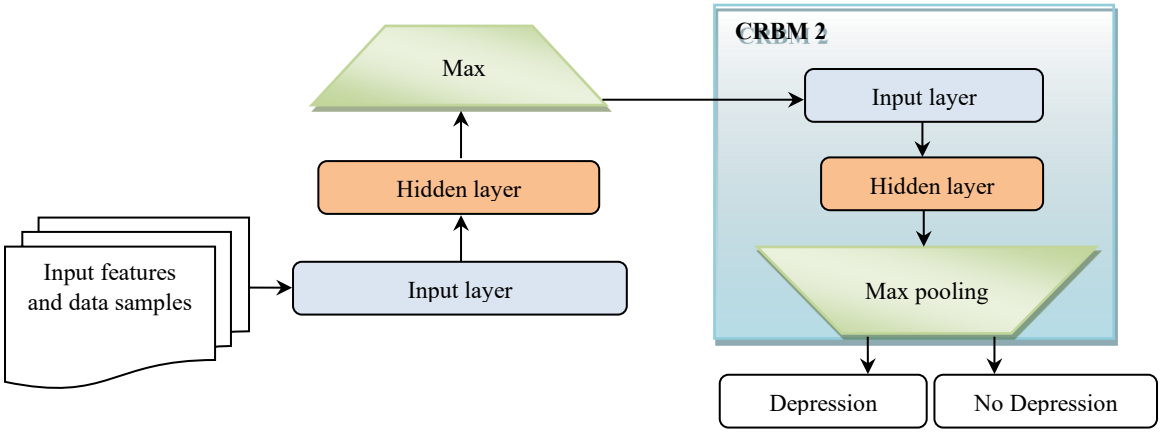

Fig. 2 Structural network of CDBN

Figure 2 exposes the structural network of a CDBN for accurate depression prediction in the social network. The learning process of CDBN consists of two major phases: layer-by-layer training and fine-tuning. In the layer-by-layer training phase, each layer of the CDBN processes weighted input data samples, and results are sent to the subsequent layer. The fine-tuning phase performs the error backpropagation to refine the hyperparameters and enhance the model's accuracy. Initially, the layer-by-layer approach is carried out with the given input data samples. In this phase, CDBM utilises the Convolutional Restricted Boltzmann Machines (CRBMs), stochastic neural networks including three layers: visible, hidden, and max-pooling, as shown in Figure 2. Each layer includes numerous neurons or nodes to receive and transfer the input samples from one layer to another.

The visible layer receives an input sample. As revealed in figure 2, the CDBNs consist of training set $\{TS, Y\}$ where TS, signifies a training tweet samples$TS = \{TS_1, TS_2, TS_3, \ldots TS_n\}$' collected from the dataset and a label or output 'Y' indicating its category, which provides the different classes such as depression or no depression. Input tweet samples are associated with weight. '$Q_1, Q_2, \ldots, Q_n$' added with bias 'b'. The neuron activation probability of the visible layer of CRBMs is formulated as given below,

$$P_v = f\left(\sum_{i=1}^{n} TS_i * Q_v\right) + b_v \quad (1)$$

Where, P_v indicates an activation probability of the neuron in the visible layer, f represents a sigmoid activation function, 'TS_i' denotes input tweet data samples, Q_v denotes weights in the visible layer, b_v indicates bias of the visible layer. If neuron activation probability $P_v = 1$, then input samples are received by a hidden layer where pre-processing is carried out.

3.2.1. Pre-Reprocessing

It ensures the dataset is clean, reliable, and suitable for analysis. In the pre-processing step, word tokenisation, stop word removal, and word stemming processes are performed. The proposed DOPR-GCDBN model utilises the Penn Treebank to split the input tweets into individual words or tokens. The Penn Treebank Tokenizer is a specific method to split the text through punctuation and spaces. The tokenisation process is expressed as follows,

$$TS_i \xrightarrow{PTT} [\text{'}W_1\text{'},\ \text{'}W_2\text{'},\ \text{'}W_3\text{'}\ \ldots\text{'}W_m\text{'}] \quad (2)$$

Where, TS_idenotes a tweet sample, W_1', 'W_2', 'W_3' ...'W_m Denotes words extracted from the tweets using Penn Treebank Tokenizer 'PTT'.After the tokenisation, the stop words deletion process is carried out to remove regular words that do not contribute significant meaning. The main advantage of the stop-word deletion process is that it reduces the size and makes it more manageable and faster. A canonical correlative stop-word removal process is employed in the proposed DOPR-GCDBN model to detect the stop words and perform the sentimental analysis. These words are often sorted out to reduce the size of the text data samples and extract the more significant, meaningful words for accurate analysis. The DOPR-GCDBN model uses the list of stop word files to analyse words in input tweet samples. The canonical correlation is a statistical method used to measure the similarity between the words in tweets and a list of stop-word files. Let us consider the number of words. 'W_1', 'W_2', 'W_3' ...'W_m in tweet samples, 'TS' and a list of stop words files. ' W_{SL}'. Similarity is expressed as follows,

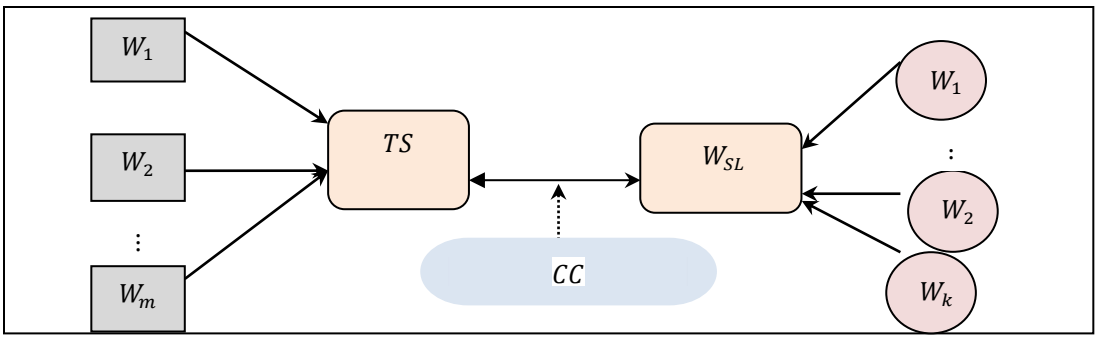

Fig. 3 Canonical correlation

Canonical correlation is measured based on covariance between words, and the mean is expressed as

$$Cov\ (W_j, W_k) = \frac{\sum(W_j - M_j)(W_k - M_k)}{m} \quad (3)$$

Where, $Cov\ (W_j, W_k)$ denotes a covariance between the words W_j and W_k , M_j denotes a mean frequency of the word. 'W_j' and M_k indicates a mean frequency of the word 'W_k'. The words with high average covariance are said to be a stop word. Otherwise, the words are identified as not a stop word. Finally, the identified stop words are removed from the tweet samples. Word stemming is the process of removing the additional words from the origin or root word. On the other hand, the word-stemming process removes the suffixes and offers the main words.

Table 3. Example of stemming process

Word extracted from tweet	Stemming	Root word
Finishing	ing	Finish
frequently	ly	frequent
Ended	ed	End

In Table 3, word ends with 'ing', 'ly', and 'ed', which are called suffixes, and it removed and obtained root word finish, frequent, and end. Pre-processing results are given to the maxpooling layer.

3.2.2. Keyword Extraction

Keyword extraction is the significant process of automatically identifying the important words within a text. The maxPooling layer is used to minimise the spatial dimensions of input samples while keeping the important information. It is applied to reduce computational complexity and extract dominant features or keywords. The proposed DOPR-GCDBN model utilises the dragonfly optimisation algorithm to extract keywords or features from the pre-processed texts, thereby minimising the time required for classification. Dragonfly Optimization is a metaheuristic optimisation technique used to handle high-dimensional optimisation problems while handling the multiple objectives functions. The most important behaviour of the dragonfly is dynamic movement and searching for its food source. Contrary to other optimisation techniques, the main objective of radial kernelised Dragonfly Optimization is also effective in handling significant data variables. It can manage and process extensive data efficiently. In this optimisation, the dragonflies are related to the number of words in the text. The food source is related to fitness functions, including objective functions. In the proposed optimisation algorithm, population initialisation refers to generating an initial set of individuals, dragonflies, or words from which the algorithm developed to find optimal solutions.

$$W_j = W_1, W_2, \dots W_b \tag{4}$$

Where, W_j denotes 'b' number of words in search space. After the population initialisation process, the fitness of each dragonfly in the current population is estimated. The radial kernel function is applied to measure the contextual appearance of the word in the given texts. The kernel function is a statistical method to measure word similarity.

$$RK = \exp\left[-\frac{|W_f - W_{avg}|^2}{2\sigma^2}\right] \tag{5}$$

$$FF = \begin{cases} RK > T \ ; \ select\ as\ keyword \\ otherwise; \quad Not\ selected \end{cases} \tag{6}$$

Where RK denotes a radial kernel, W_f stands for the contextual appearance frequency of the word 'W' in a given text and W_{avg} denotes an average appearance frequency of a word in a given text, σ indicates a deviation, FF denotes a fitness function, Tdenotes threshold. If the contextual appearance frequency of a word in the given text is lesser than its average frequency, the RK value becomes smaller than the threshold, and the word is not selected as a keyword. On the other hand, if the contextual appearance frequency of a word in the given text is higher than its average frequency, the RK value exceeds the threshold, and the word is selected as a keyword. This approach helps select relevant keywords for the classification process while removing less significant words. Based on the fitness estimation, different swarming behaviours of each dragonfly are determined in the search space. The behaviours of the dragonflies are separation, alignment, cohesion, and attraction towards the food source. These behaviours help discover the optimal global solution for the given population. Initially, the separation behaviour is executed to find the current position of each dragonfly and their neighbouring position.

$$SE_W = -\left|\left(X_i - X_g\right)\right| \tag{7}$$

Where, SE_W denotes the separation behaviour of dragonflies, $X_i$stands for the current position of a dragonfly, X_g denotes the neighbouring position of dragonflies. The second behaviour of dragonflies is alignment, which measures the movement velocity of dragonflies in the direction of neighbouring dragonflies.

$$AL_W = \frac{1}{r}\sum_{g=1}^{r} V_g \tag{8}$$

Where,AL_W refers to an alignment behaviour, V_g denotes a velocity of 'neighbouring dragonflies, $'r$' indicates many neighbouring dragonflies. The third behaviour is cohesion, which determines the movement of dragonflies towards the centre of their neighbourhood.

$$Ch_w = \frac{1}{r}\sum_{g=1}^{r}[X_g - X_i] \tag{9}$$

Where, Ch_w signifies a cohesion process of the dragonfly,X_g denotes the position of the neighbouring dragonfly, X_i Indicates a position of a current dragonfly, r is the number of neighbourhoods. Finally, the attraction towards the food source is executed based on the current position of the food source and the position of the dragonfly expressed as follows.

$$At_W = |X_{Food} - X_i| \tag{10}$$

Where, At_Wrepresents an attraction towards a food source, X_{Food} indicates the position of the food source, X_i denotes the current position of a dragonfly. Based on the above-said parameters, the position of the dragonfly gets updated according to their neighbourhoods,

$$X^{new} = X_i + \nabla X_R \quad (11)$$

$$\nabla X_R = \{\alpha_1 . SE_W + \alpha_2 . AL_W + \alpha_3 . Ch_w + \alpha_4 . At_W\} + \varphi * \nabla X_i \quad (12)$$

Where X^{new} refers to an updated position of the dragonfly, $X_i$indicates the current position of a dragonfly, ∇X_Rindicates step vector to find the movement direction of the dragonfly,α_1 indicates the weight of the separation function (SE_W), α_2 indicates the weight of the alignment function. 'AL_W', $\alpha_3$denotes the weight of cohesion Ch_w,α_4 denotes the weight of attraction towards a food source At_W, previous step vector. '∇X_i', φ denotes inertia weight. This process is continued until it reaches the maximum iterations. At the end of the maximum iterations reached, optimal keywords are selected.

3.2.3. Classification

The depression prediction process is done through classification with selected optimal keywords to improve accuracy. A Piecewise regression algorithm is employed in DOPR-GCDBN to identify the depression and non-depression texts. Piecewise regression is a machine learning technique used to determine the relationship between optimal keywords in the training and testing sets, then provides the final classification outcomes by setting a threshold value. Russel–Rao similarity 'R' is applied between the keywords as follows,

$$R = \frac{W_o \cap W_t}{M} \quad (13)$$

Where, W_o the optimal number of keywords in the training set, W_t denotes keywords in the testing set, M denotes the total number of keywords in the testing set.

$$Y = \begin{cases} 1; & depression \ \ for \ R > T \\ 0; & no \ depression \ \ for \ R < T \end{cases} \quad (14)$$

Where Y denotes a Piecewise regression outcome, T indicates a threshold, R denotes a Russel–Rao similarity function. If the regression outcome 'Y' returns '1', the tweet samples are classified as depression. Otherwise, the tweets samples are classified as no depression. Accurate detection of the tweet samples is classified as depression and no depression. After classification, the fine-tuning process is executed. In the fine-tuning process, the error rate is measured based on a squared difference between actual as well as predicted classification results,

$$ERR = [Y_{act} - Y_{Pre}]^2 \quad (15)$$

Where 'ERR' represents the error rate,Y_{act}indicates the actual classification output, Y_{Pre} Symbolises the predicted classification output. In order to reduce error, the proposed technique utilises the adaptive stochastic gradient method to adjust the weight between the layers.

$$Q^{new} = Q^{old} - \eta \left[\frac{\partial ERR}{\partial Q^{old}}\right] \quad (16)$$

Where, Q^{new} indicates new weight, Q^{old}designates a current weight, η represents learning rate, $\frac{\partial ERR}{\partial Q^{old}}$ symbolises first-order derivative algorithm to find out the local minimum of a function by adjusting the current weight 'Q^{old}'.

This process is continued until the error gets minimised. Optimal weight values are selected to minimise the classification error. Depression prediction outcomes are determined by minimising the classification error at the output layer.

//Algorithm 1: Dragonfly Optimised Piecewise Regression-based Gradient Convolutional Deep Belief Network
Input: Datasets 'Ds', tweet samples TS = { $TS_1, TS_2, TS_3, \ldots TS_n$}   Output: Improve depression prediction Accuracy
Begin   Step 1: Collect the tweet samples TS = { $TS_1, TS_2, TS_3, \ldots TS_n$} from dataset   Step 2: Input the samples 'TS' given to the visible layer   Step 3: For each sample 'TS'   Step 4: Formulate the neuron activation probability using (1)   Step 5:Apply Penn Treebank Word Tokenizer to obtain the words W_1', 'W_2', 'W_3' ... 'W_m Step 6: Apply canonical correlation to remove the stop words using (3)   Step 7: Perform word stemming   Step 8: End For   Step 9: For each pre-processed results   Step 10: Initialize dragonflies' population of words $W_j = W_1, W_2, \ldots W_b$ Step 11: for each dragonflies   Step 12: Measure the radial kernel using (5)   Step 13: Measure fitness using (6)   Step 14: End for   Step 15: End for   Step 16: for each dragonfly

Step 17: Calculate four behavior using (7) (8) (9) (10)
Step 18: End for
Step 19: While ($t \leq Max\_t$)do
Step 20: Update the position of dragonflies based on Equations (11)
Step 21: Increment $t = t + 1$
Step 22:Go to step 19 until it converges
Step 23: End while
Step 24: Return (optimal keywords)
Step 25: For each extracted optimal keyword in the training set
Step 26: For each extracted optimal keyword in the testing set
Step 27: Measure Russel–Rao similarity using (13)
Step 28: End for
Step 29: End for
Step 30:If ($Y = 1$) then
Step 31: tweet samples are classified as 'depression'
Step 32: else
Step 33: tweet samples are classified as ' no depression'
Step 34: End if
Step 35: End for
Step 36: For each classification outcomes
Step 37: Compute the error rate using (15)
Step 38: Update the weight using (16)
Step 39: End for
Step 40:Obtain the optimal weight
Step 41: Obtain the final classification results at the output layer
End

4. Experimental Setup

Experimental evaluation of DOPR-GCDBN model and CBA [1] and BERT-CNN [2] are implemented in Python language using Depression: Twitter Dataset + Feature Extraction.

5. Performance Comparative Analysis

5.1. Evaluation Metrics Analysis

5.1.1. Depression Prediction Accuracy

It refers to the proportion of correctly identified depression and no depression tweet samples out of the total number of tweet samples and calculated as,

$$DA = \left(\frac{Tps+Tng}{Tps+Tng+Fps+Fng}\right) * 100 \quad (17)$$

5.1.2. Precision

It refers to the ratio of detecting the depression and no depression tweet samples, and The precision is mathematically computed as,

$$PRC = \left(\frac{Tps}{Tps+Fps}\right) \quad (18)$$

PRC denotes a precision, Tpsdenotes the true positive, Fps represents the false positive.

5.1.3. Recall

It refers to the ratio of depression and no depression-tweet samples determined by the total number of tweet samples in the data set. It is known as sensitivity and expressed as,

$$REC = \left(\frac{TP}{TP+FN}\right) \quad (19)$$

5.1.4. F1 Score

It is also known as the F-measure classification performance metric to balance the mean of precision as well as recall, and it is formulated as follows,

$$F1\_Score = 2 * \left(\frac{PRE*REC}{PRE*REC}\right) \quad (20)$$

5.1.5. Specificity

Specificity in Depression prediction measures algorithm accurately identifies the data samples that did not have depression as truly negative cases. Specificity is expressed as follows,

$$SPE = \left(\frac{TN}{TN*FP}\right) \quad (21)$$

5.1.6. Depression Prediction Time

An evaluation metric measured based on the time the algorithm consumes to detect depression, with no depression tweet samples.

It is calculated as

$$\boldsymbol{DPT = \sum_{i=1}^{n} TS_i * TM(DD)} \quad (22)$$

Table 4. Depression prediction accuracy versus the number of tweet samples

Number of tweet samples	DOPR-GCDBN	CBA	BERT-CNN
2000	97	94	92.5
4000	96.89	93.56	92.05
6000	97.56	93.78	92
8000	97.33	94.05	91.56
10000	97.05	93.12	91.11
12000	97.41	93.45	91.06
14000	96.78	93.11	91
16000	96.05	93.1	90.56
18000	97.45	93.08	90.33
20000	96.74	93	90.22

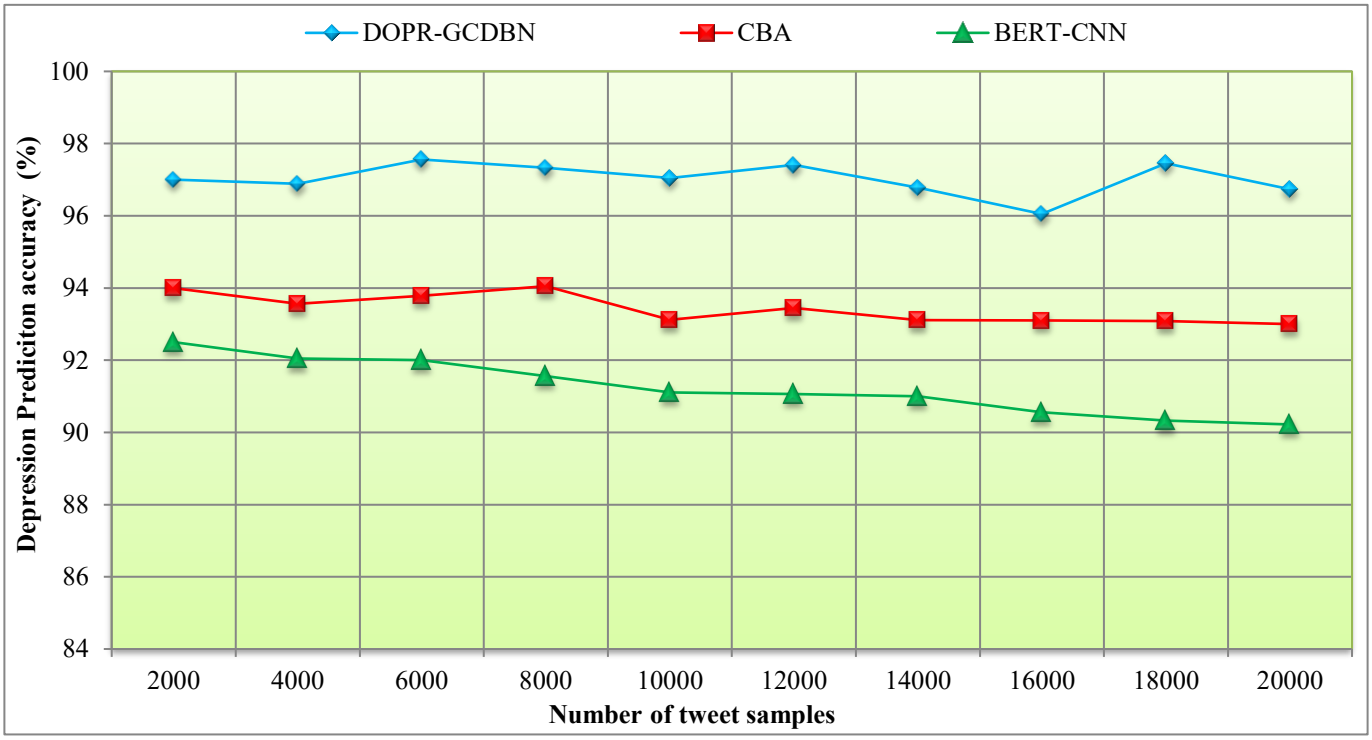

Fig. 4 Graphical illustration of depression prediction accuracy

Figure 4 depicts the accuracy of depression prediction using three methods: the DOPR-GCDBN model CBA [1] and BERT-CNN [2]. The horizontal axis indicates the number of tweet samples, ranging from 2000 to 20000, while the vertical axis represents the accuracy of the depression prediction. The DOPR-GCDBN model reveals better accuracy than the other approaches. For example, in the first iteration with 2000 samples, the DOPR-GCDBN model achieved a depression prediction accuracy of 97%.

The accuracy of existing models [1, 2] were observed to be 94% and 92.5%, respectively. Ten results were observed for each method by varying the number of tweet samples, allowing a comprehensive comparison of their performance. The comparative analysis shows that the DOPR-GCDBN model outperformed [1, 2] by 4% and 6%, respectively. This improvement is accomplished due to the proposed convolutional deep belief network, which analyses keywords using a Piecewise regression.

The regression technique determines the relationship between the optimal keywords in the training set and those in the testing set. It provides the final classification outcomes based on the similarity between Russel and Rao.

Furthermore, the fine-tuning process of the DOPR-GCDBN model, utilising the adaptive gradient method, reduces the classification error rate.

This result minimises the false positives and negatives while considerably enhancing the true positive and true negative rates. As a result, the accuracy of the depression prediction is greatly enhanced by applying the DOPR-GCDBN model compared to the other methods.

Table 5. Precision versus number of tweet samples

Number of Tweet Samples	DOPR-GCDBN	CBA	BERT-CNN
2000	0.962	0.933	0.923
4000	0.965	0.931	0.915
6000	0.966	0.928	0.905
8000	0.958	0.932	0.912
10000	0.963	0.937	0.917
12000	0.967	0.939	0.916
14000	0.959	0.937	0.911
16000	0.963	0.927	0.907
18000	0.958	0.93	0.917
20000	0.964	0.928	0.913

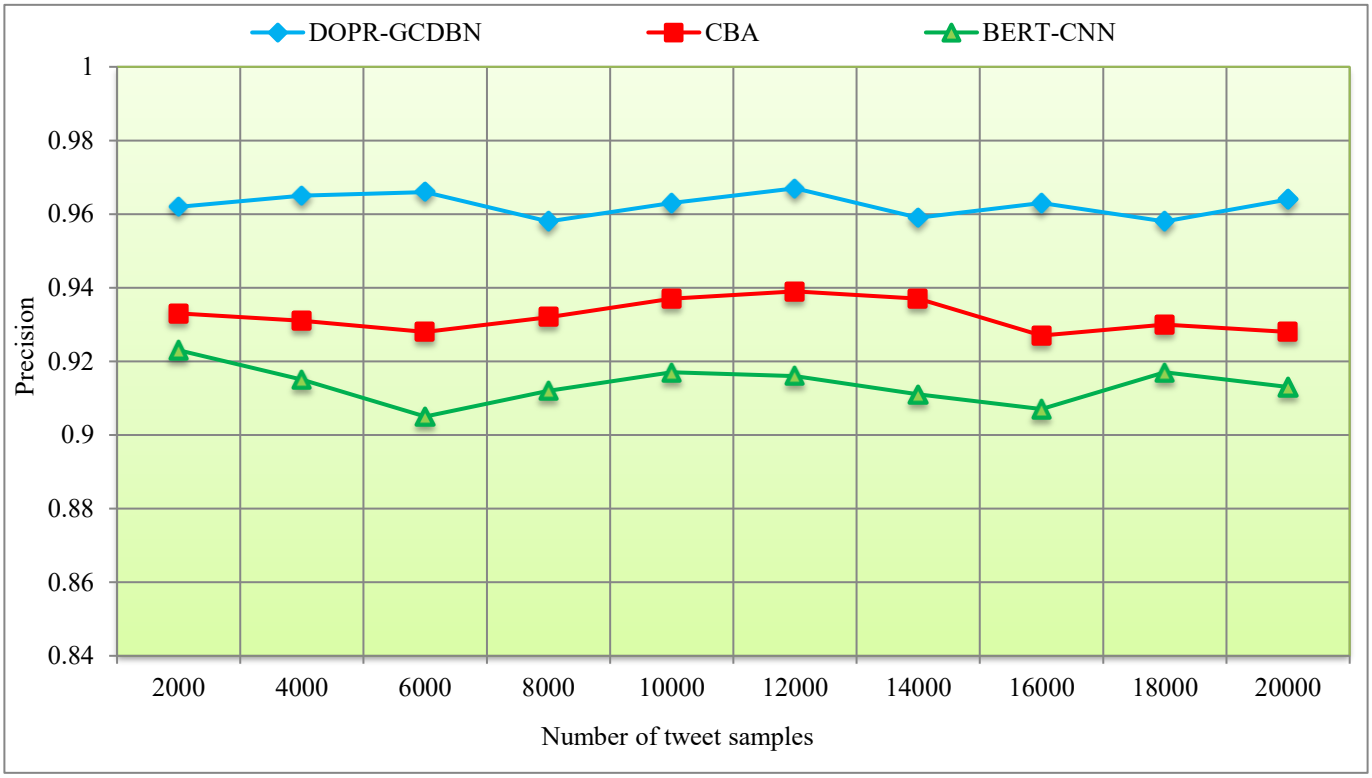

Fig. 5 Graphical illustration of precision

Figure 5 shows the graphical analysis of precision in the depression prediction with the number of tweet data samples ranging from 2000 to 20000. Three methods were applied, namely the DOPR-GCDBN model, CBA [1], and BERT-CNN [2], to evaluate precision during depression prediction from the twitter dataset.

As shown in Figure 5, the 'x' axis represents the tweet data samples count, while the 'y' axis indicates the performance of the precision. The experimental results conclude that the DOPR-GCDBN model achieved better precision than the other two approaches. Let us consider 2000 tweet samples from the dataset to compute the precision. The DOPR-GCDBN model observed a precision of 0.962, and performances of precision were observed to be 0.933 and 0.923, respectively. Each method's performance results were observed with varying Twitter data samples. The DOPR-GCDBN model's observed results are compared to the existing methods.

The overall comparative analysis designates that the DOPR-GCDBN model improved the precision by 3% to [1] and 5% to [2], respectively. This improved performance is achieved by utilising a convolutional Deep Belief Network for analysing training and testing keywords through the regression analysis, which provides a better actual positive rate.

Furthermore, integrating the adaptive gradient method reduces classification errors, thereby improving the accuracy of depression prediction by minimising false positives. Therefore, the DOPR-GCDBN model outperforms existing methods by achieving higher precision in depression detection.

Table 6. Recall versus number of tweet samples

Number of Tweet Samples	DOPR-GCDBN	CBA	BERT-CNN
2000	0.980	0.951	0.932
4000	0.975	0.948	0.931
6000	0.972	0.944	0.928
8000	0.978	0.943	0.926
10000	0.977	0.94	0.929
12000	0.981	0.947	0.927
14000	0.975	0.943	0.925
16000	0.982	0.942	0.92
18000	0.976	0.945	0.923
20000	0.979	0.938	0.92

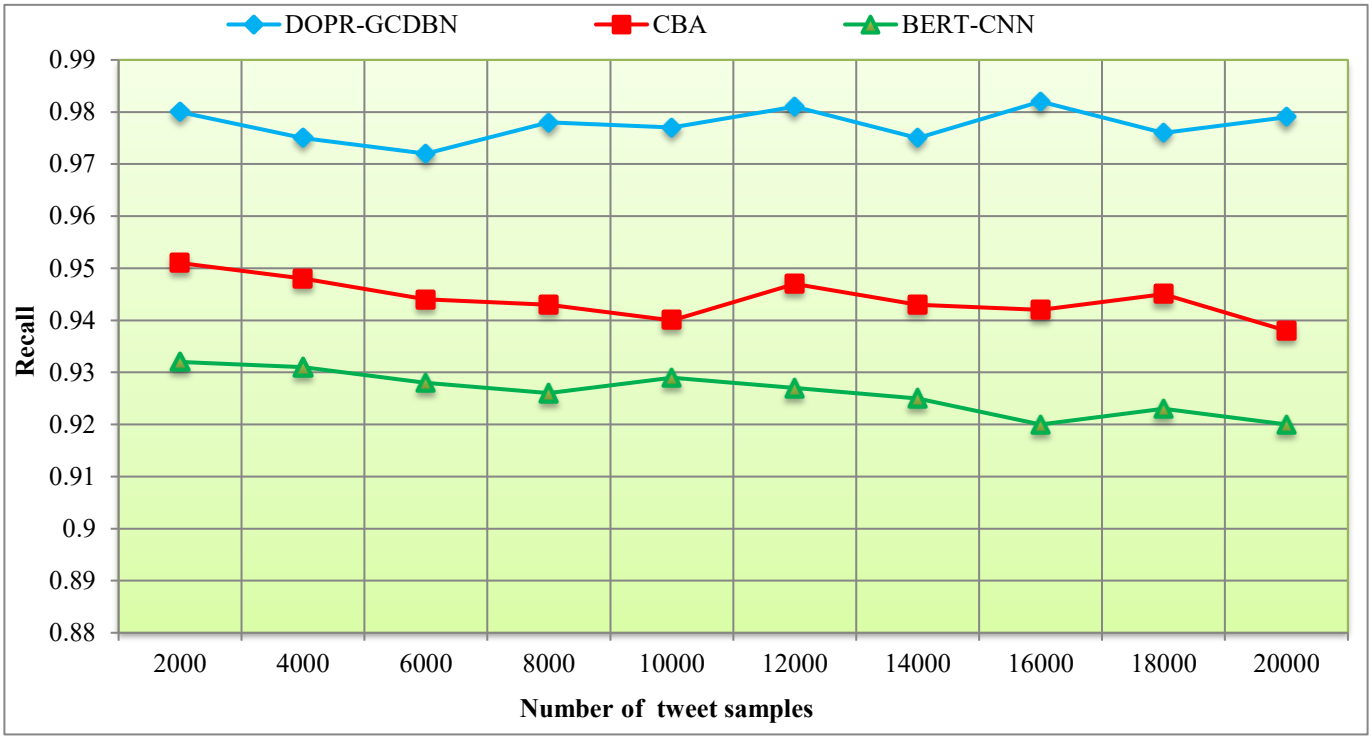

Fig. 6 Graphical illustration of recall

Figure 6 provides a graphical representation of the graphical results of recall concerning the number of tweet samples collected from 2000 to 20000 using three methods: the DOPR-GCDBN model, CBA [1], and BERT-CNN [2]. The horizontal axis indicates the number of tweet samples, while the vertical axis symbolises recall performance. The DOPR-GCDBN model exposes comparatively better results in achieving recall than [1, 2]. For the first run, with 2000 tweet samples, the DOPR-GCDBN model achieved a recall performance of 0.980, while the existing methods [1, 2] were achieved to be 0.951 and 0.932, respectively. Comparing the DOPR-GCDBN model with the existing methods, the recall performance is improved by 4% and 6% compared to [1] and [2], respectively. The proposed convolutional deep belief network model decreases the squared difference between the actual and predicted outputs by applying the stochastic gradient method to find the optimal weights for training the layers. This process is iterated until we find a minimal error, thereby reducing the false-negative rates and increasing the actual positive rate in depression detection.

Table 7. F1 score versus the number of tweet samples

Number of tweet samples	DOPR-GCDBN	CBA	BERT-CNN
2000	0.970	0.941	0.927
4000	0.969	0.939	0.922
6000	0.968	0.935	0.916
8000	0.967	0.937	0.918
10000	0.969	0.938	0.922
12000	0.973	0.942	0.921
14000	0.966	0.939	0.917
16000	0.972	0.934	0.913
18000	0.966	0.937	0.919
20000	0.971	0.932	0.916

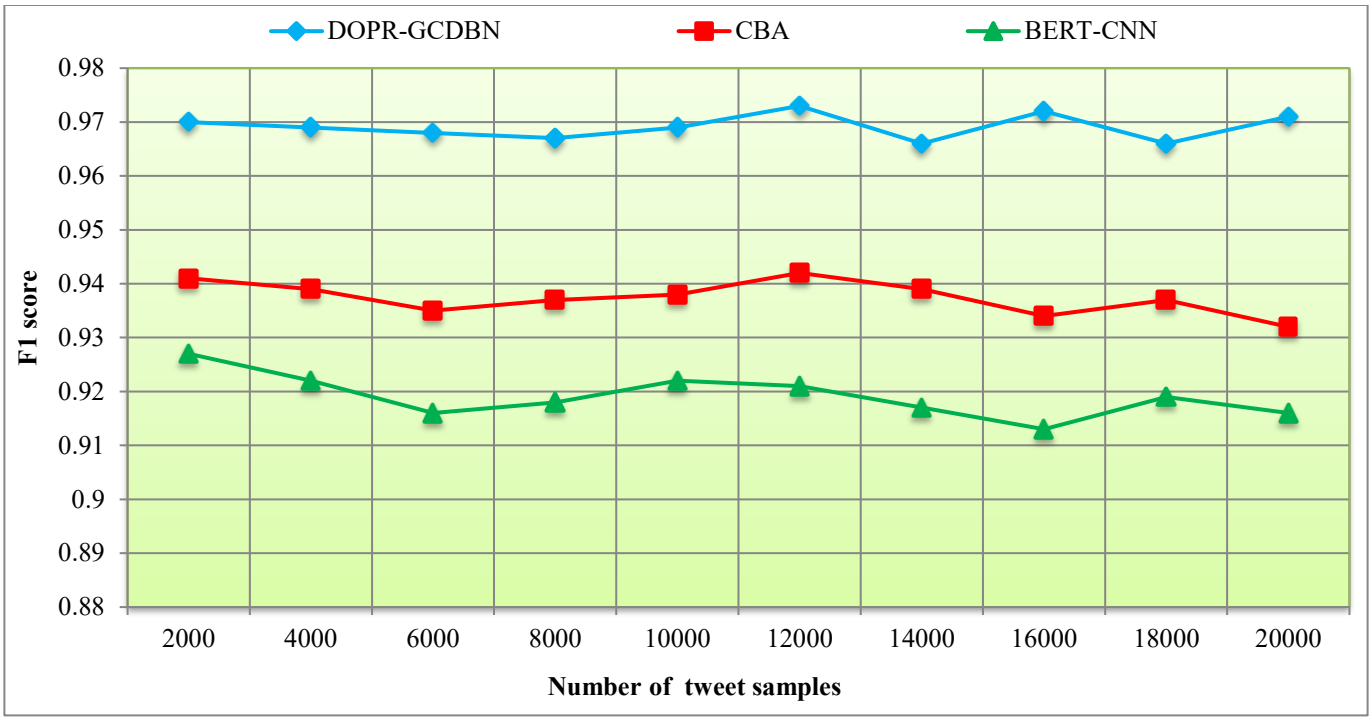

Fig. 7 Graphical illustration of F1 score

Figure 7 illustrates the experimental results of the F1-score versus number of tweet samples collected from the dataset. The observed results designate that the proposed DOPR-GCDBN model significantly improved F1-score performance. For example, in 2000 samples taken in the first run, the DOPR-GCDBN model [1, 2] achieved an F1-score of 0.970, 0.941 and 0.927, respectively. This examination results underline that the DOPR-GCDBN model achieved better performance in depression detection. The average of ten comparison results indicates that the DOPR-GCDBN model improved the F1-score by 3% and 5% compared to [1, 2], respectively. This improvement is achieved because the DOPR-GCDBN model increases the precision and recall performance in depression detection.

The performance outcomes of the specificity during the depression prediction using three methods, namely the DOPR-GCDBN model, CBA [1], and BERT-CNN [2], are illustrated in Figure 8. The overall specificity performance outcomes were estimated based on many tweet samples. It is obvious from the observed results that the DOPR-GCDBN model outperforms the existing methods in terms of achieving better specificity. This is because it is a practical application of the convolutional deep belief classifier, leading to better specificity. Overall, the performance results reveal that the specificity of the DOPR-GCDBN model is enhanced by 3% compared to [1] and by 6% compared to [2]. The performance results of the depression prediction time using three different methods, the DOPR-GCDBN model, CBA [1] and BERT-CNN [2], are shown in Figure 9. As tweet samples increase, the depression prediction time also increases. Therefore, the depression prediction time is proportional to the number of tweet samples. The observed results specify that the performance of the DOPR-GCDBN model is relatively better in achieving minimal time consumption than the existing methods. Considering 2000 tweet samples, the time taken to perform depression prediction was found to be '$44ms$'. However, the time consumption of existing [1, 2] was found to be $56ms$' and $64ms$. The examined results show that the DOPR-GCDBN model reduces the depression prediction time. After obtaining the ten results, the overall time consumption of the DOPR-GCDBN model is compared to the existing results. The average of ten results shows that the proposed DOPR-GCDBN model minimised the depression prediction time consumption by 12% and 24% compared to the [1, 2], respectively.

Table 8. Specificity versus number of tweet samples

Number of Tweet Samples	DOPR-GCDBN	CBA	BERT-CNN
2000	0.958	0.927	0.917
4000	0.952	0.925	0.911
6000	0.948	0.922	0.905
8000	0.956	0.92	0.907
10000	0.955	0.911	0.908
12000	0.953	0.918	0.895
14000	0.95	0.917	0.888
16000	0.957	0.92	0.904
18000	0.949	0.927	0.895
20000	0.954	0.925	0.904

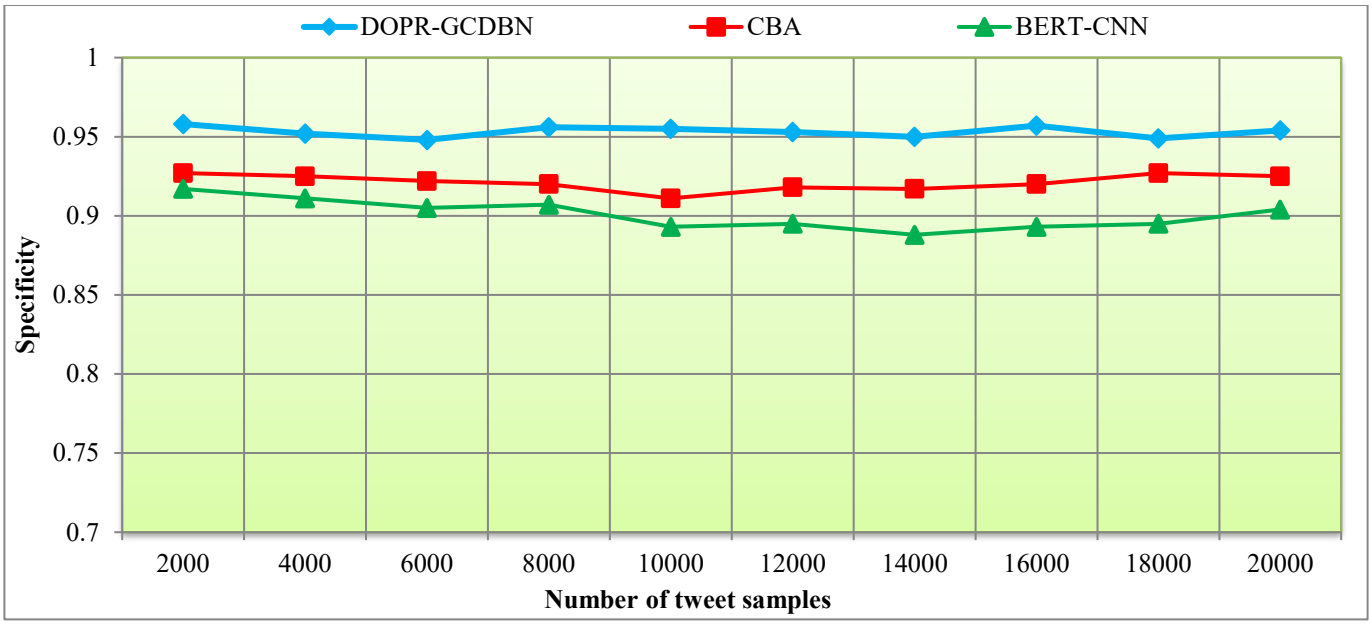

Fig. 8 Graphical illustration of specificity

Table 9. Depression prediction time versus the number of tweet samples

Number of Tweet Samples	DOPR-GCDBN	CBA	BERT-CNN
2000	44	56	64
4000	52.6	60.3	68.6
6000	58	65	70.5
8000	65.7	73.6	77.6
10000	72	85.5	88.7
12000	83	89	92.6
14000	91.6	95.6	109.4
16000	95.7	105.7	117.6
18000	105.6	113.5	123.8
20000	112.5	126.6	147.5

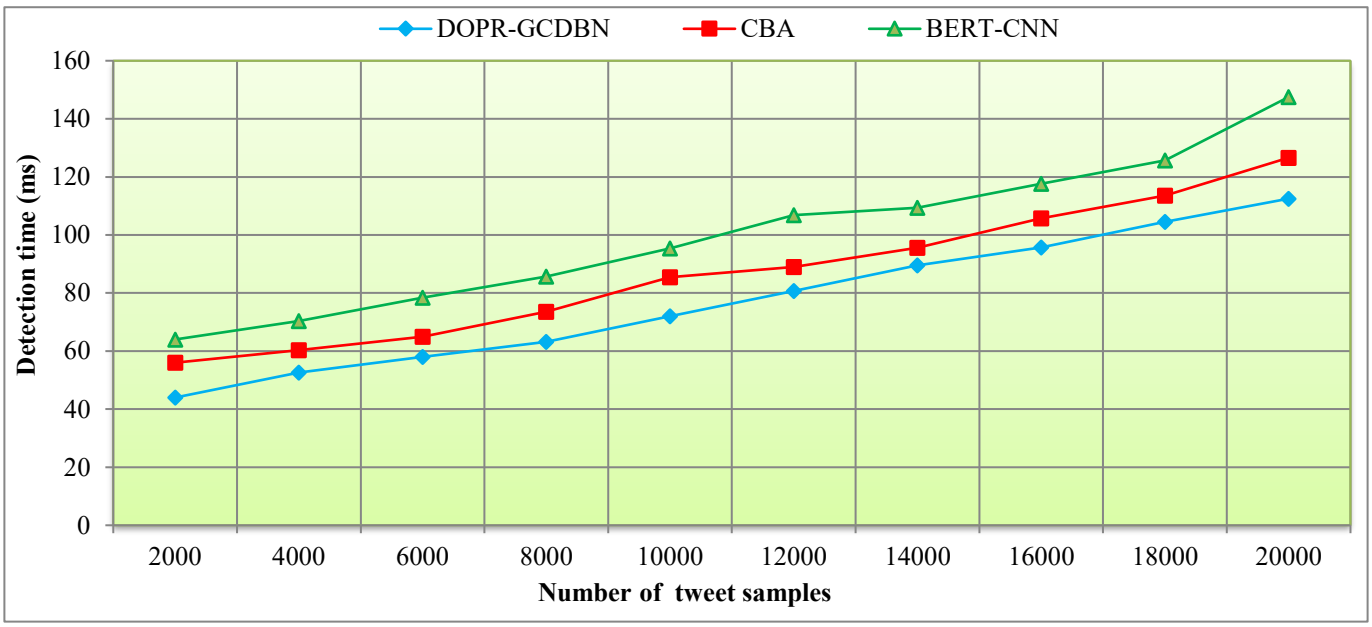

Fig. 9 Graphical illustration of depression prediction time

This is because the DOPR-GCDBN model integrates the pre-processing keyword extraction within the layer of deep learning architecture, which minimises the depression prediction time. Moreover, Tokenizer has applied to partition the review into several words. After that, the canonical correlation method is applied to remove the stop words. The Dragonfly optimisation is applied to select the important keywords from the tweets to perform the classification, resulting in minimised time consumption and improved accuracy of depression detection.

6. Discussion

This study compares the proposed DOPR-GCDBN with the existing CBA [1], and BERT-CNN [2] is discussed with Depression: Twitter Dataset + Feature Extraction (https://www.kaggle.com/datasets/infamouscoder/mentalhealth-social-media) based on various parameters, such as depression prediction accuracy, precision, recall, F1 score and detection time. The results confirm that the proposed DOPR-GCDBN method improved accuracy by 5%, precision by 4% and recall by 5%, F1-score by 4%, specificity by 5%, depression prediction time by 18% when compared to the existing methods [1, 2] using the Depression dataset.

7. Conclusion

Mental health is a vital part of human health and well-being that facilitates people's management of the stresses of life and so on. A novel deep learning model called DOPR-GCDBN has been developed to enhance depression prediction accuracy while minimising time consumption error.

References

[1] Joel Philip Thekkekara, Sira Yongchareon, and Veronica Liesaputra, "An Attention-Based CNN-BiLSTM Model for Depression Prediction on Social Media Text," *Expert Systems with Applications*, vol. 249, part c, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[2] Cao Xin, and Lailatul Qadri Zakaria, "Integrating Bert with CNN and BiLSTM for Explainable Detection of Depression in Social Media Contents," *IEEE Access*, vol. 12, pp. 161203-161212, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[3] Nawshad Farruque et al., "Depression Symptoms Modelling from Social Media Text: An LLM Driven Semi-Supervised Learning Approach," *Language Resources and Evaluation*, vol. 58, pp. 1013-1041, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[4] Kathiresh Murugesan et al., "Homomorphic Encryption, Privacy-Preserving Feature Extraction, and Decentralized Architecture for Enhancing Privacy in Voice Authentication," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 15, no. 2, pp. 2150-2160, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[5] Yongfeng Tao et al., "DepMSTAT: Multimodal Spatio-Temporal Attentional Transformer for Depression Detection," *IEEE Transactions on Knowledge and Data Engineering*, vol. 36, no. 7, pp. 2956-2966, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[6] Muhammad Asad Abbas et al., "Novel Transformer Based Contextualized Embedding and Probabilistic Features for Depression Prediction from Social Media," *IEEE Access*, vol. 12, pp. 54087-54100, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[7] Wenli Zhang et al., "Depression Prediction Using Digital Traces on Social Media: A Knowledge-aware Deep Learning Approach," *Journal of Management Information Systems*, vol. 41, no. 2, pp. 546-580, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[8] Muthukumar Palani, and Velumani Thiyagarajan, "An Intelligent Early Detection of Melanoma Using Fuzzy Neural Networks with Java Optimization (FNN-JO) Classifier," *International Journal of Engineering Trends and Technology*, vol. 72, no. 7, pp. 75-82, 2024. [CrossRef] [Publisher Link]

[9] Marios Kerasiotis, Loukas Ilias, and Dimitris Askounis, "Depression Detection in Social Media Posts Using Transformer-Based Models and Auxiliary Featuresa," *Social Network Analysis and Mining*, vol. 14, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[10] Tasnim Ahmed et al., "Decoding Depression: Analyzing Social Network Insights for Depression Severity Assessment with Transformers and Explainable AI," *Natural Language Processing Journal*, vol. 7, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[11] Rafael Salas-Zárate et al., "Mental-Health: An NLP-Based System for Detecting Depression Levels through User Comments on Twitter (X)," *Mathematics*, vol. 12, no. 13, pp. 1-30, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[12] Sahira Vilakkumadathil, and Velumani Thiyagarajan, "Exploring Diverse Prediction Models in Intelligent Traffic Control," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 38, no. 1, pp. 393-402, 2025. [CrossRef] [Publisher Link]

[13] Felix Indra Kurniadi et al., "BERT and RoBERTa Models for Enhanced Detection of Depression in Social Media Text," *Procedia Computer Science,* vol. 245, pp. 202-209, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[14] Bayode Ogunleye, Hemlata Sharma, and Olamilekan Shobayo, "Sentiment Informed Sentence BERT-Ensemble Algorithm for Depression Detection," *Big Data and Cognitive Computing*, vol. 8, no. 9, pp. 1-16, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[15] Filza Rehmani et al., "Depression Prediction with Machine Learning of Structural and Non-Structural Dual Languages," *Healthcare Technology Letters*, vol. 11, no. 4, pp. 218-226, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[16] Ramaraj Muniappan et al., "Optimization Techniques Applied on Image Segmentation Process by Prediction of Data Using Data Mining Techniques," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 15, no. 2, pp. 2161-2171, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[17] Meena Kumari, Gurpreet Singh, and Sagar Dhanraj Pande, "Depressonify: BERT a Deep Learning Approach of Detection of Depression," *EAI Endorsed Transactions on Pervasive Health and Technology*, vol. 10, 2024. [Google Scholar] [Publisher Link]

[18] Samar Samir Khalil, Noha S. Tawfik, and Marco Spruit, "Federated Learning for Privacy-Preserving Depression Prediction with Multilingual Language Models in Social Media Posts," *Patterns*, vol. 5, no. 7, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[19] Yi Xiao et Al., "Empirical Insights into the Interaction Effects of Groups at High Risk of Depression on Online Social Platforms with NLP-Based Sentiment Analysis," *Data and Information Management*, vol. 8, no. 4, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[20] Velumani Thiyagarajan, "Adaptive Feature Learning for Robust Pathogen Detection in Plants," *2024 4th International Conference Adaptive Feature Learning for Robust Pathogen Detection in Plants on Sustainable Expert Systems (ICSES)*, Kaski, Nepal, pp. 1347-1353, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[21] Rula Kamil, and Ayad R. Abba, "Predicating Depression on Twitter Using Hybrid Model BiLSTM-XGBOOST," *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 6, pp. 3620-3627, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[22] Ozay Ezerceli, and Rahim Dehkharghani, "Mental Disorder and Suicidal Ideation Detection from Social Media Using Deep Neural Networks," *Journal of Computational Social Science*, vol. 7, pp. 2277-2307, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[23] Seyedeh Leili Mirtaheri, Sergio Greco, and Reza Shahbazian, "A Self-Attention TCN-Based Model for Suicidal Ideation Detection from Social Media Posts," *Expert Systems with Applications*, vol. 255, part d, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[24] Eliseo Bao, Anxo Pérez, and Javier Parapar, "Explainable Depression Symptom Detection in Social Media," *Health Information Science and Systems*, vol. 12, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[25] Sayani Ghosal, and Amita Jain, "Depression and Suicide Risk Detection on Social Media Using Fast Text Embedding and XGBoost Classifier," *Procedia Computer Science*, vol. 218, pp. 1631-1639, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[26] Abu Bakar Siddiqur Rahman et al., "DepressionEmo: A Novel Dataset for Multilabel Classification of Depression Emotions," *Journal of Affective Disorders*, vol. 366, pp. 445-458, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[27] Shaunak Inamdar et al., "Machine Learning Driven Mental Stress Detection on Reddit Posts Using Natural Language Processing," *Human-Centric Intelligent Systems*, vol. 3, pp. 80-91, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[28] Rina Carines Cabral et al., "MM-EMOG: Multi-Label Emotion Graph Representation for Mental Health Classification on Social Media," *Robotics*, vol. 13, no. 3, pp. 1-15, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[29] Hamad Zogan et al., "Hierarchical Convolutional Attention Network for Depression Detection on Social Media and Its Impact during Pandemic," *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 4, pp. 1815-1823, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[30] Mohsinul Kabir et al., "DEPTWEET: A Typology for Social Media Texts to Detect Depression Severities," *Computers in Human Behavior*, vol. 139, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[31] S. Abirami, and P. Thirugnanam, "Depression Scale Analysis by Machine in the Field of Artificial Intelligence," *International Journal of Engineering Trends and Technology (IJETT)*, vol. 59, no. 3, pp. 130-134, 2018. [CrossRef] [Publisher Link]

[32] Shaikat Chandra Paul et al., "An Approach to Detect Suicidal Bengali Posts from Social Media Using Machine Learning Algorithms," *International Journal of Engineering Trends and Technology,* vol. 72, no. 5, pp. 43-50, 2024. [CrossRef] [Google Scholar] [Publisher Link]