Depression Prediction in Social Networks Using Whale Optimization Algorithm based Convolution Neural Networks (WOA-CNN)

N. Nalini[a*], Dr.S. Rathi[b]

Abstract
Depression is a major problem that has many different effects on individuals. Many treatments are available to help those who are depressed, but the issue is predicting those who aren't even aware that they are depressed. Furthermore, it is a leading cause of suicide ideation and significantly impairs everyday functioning. Emotion artificial intelligence is a subject of continuing study in emotion detection, particularly in the areas of social media text mining. Datasets derived from social networks are useful in a variety of subjects, including sociology and psychology. However, technical assistance is insufficient, and particular techniques are needed immediately. It encourages the development of deep learning systems for psychology to identify depressed users on social networking platforms. This paper's objective is to propose a sentiment analysis approach based on terminology and man-made criteria for calculating depressive inclination. Next, a depression detection model is built using the Whale Optimization Algorithm based Convolution Neural Networks (WOA-CNN), which takes into account parameter optimization by WOA. WOA algorithm is introduced to optimize the hyperparameters for training a CNN classifier and their layers capably. In ending, the proposed method honors social network users who have developed high-quality online mental health monitoring solutions. The results of proposed WOA-CNN classifier and existing classifiers are measured regarding accuracy, F-measure, recall, precision. Proposed technique outperforms than the state-of-the-art techniques.
Keywords: Convolutional Neural Network (CNN), Whale Optimization Algorithm (WOA), Depression, Preprocessing, Optimization, and Social Media.

1. Introduction

Public health struggles with depression, which is one of the factors contributing to suicide and is usually associated with disabilities.Early detection of the problem is essential. One of the most intricate organs in the body is the human brain. As a result, comprehending its intricacy is extremely challenging (Fleuren & Alkema, 2015). Because depression is classified as a mental condition, predicting depression is a difficult task. According to a psychiatrist, the diagnosis and treatment of depression are primarily accomplished via the use of questions and answers, as well as the use of numerous psychometric tests and theories, as well as the observation of the patient's response to them (Kumar & Ravi, 2016). However, recent research suggests that there are additional strategies for predicting depression. Patients who are depressed are hard to treat because they only focus on negative things and become trapped in a vicious cycle. As a result, depression must be avoided. People must identify their mental health in everyday life to avoid depression. They, on the other hand, are unable to detect their mental health (Lim, Tucker, & Kumara, 2017).

Twitter, Inc., Facebook, Inc., and Reddit, Inc. have become commonplace in everyday lives as platforms for sharing ideas, sentiments, and general emotional states. As a result, these platforms have evolved into important data banks for marketers and academics, who may employ user metrics, shared material, and linked data to discover preferences and tastes, as well as other attitudes and behaviors (Yadav, 2014). In reality, due to their support and capacity to comprehend someone's perspective while maintaining a comfortable emotional distance, social networks are used by patients to communicate with peers. Reddit, Inc., for

[a]Assistant professor/CSE, Sri Shakthi Institute of Engineering and Technology, Coimbatore-62. E-Mail: nalini08cse@gmail.com
[b]Professor/CSE, Government College of Technology, Coimbatore- 13. E-Mail: rathi@gct.ac.in
**Corresponding author: nalini08cse@gmail.com*

instance, is an open-source platform where members of the community may upload material and vote on it. With a significant history of prior uploads spanning several years (Sánchez et al., 2015), content entries are organized into areas of interest (signified as subreddits). It is used in thinking expression as a means of mental release and social interaction via digital means. Through the usage of social media, this has resulted in the generation of an enormous quantity of data, which is then characterised in a variety of various formats. Postings made by users of social media may include people suffering from depression, and the information gleaned from these patients' profiles might be utilised to learn more about a variety of medical conditions or characteristics of mental illness.

Researchers used a variety of techniques, including data mining, deep learning, and machine learning, to analyse social media to identify depression-related behaviours. These methods has been introduced by applying text processing methods. The four main phases in preprocessing are tokenization, stop-word elimination, transforming cases, and stemming. The process of tokenization separates the text of a document into tokens. English stopwords are removed from a text using RapidMiner's built-in stop-word collection, which eliminates every token that matches a stop-word (Wakefield, Schmitz, & Baer, 2010). In a document, all characters' cases are lowered with the convert cases command. The Porter stemming method is used to shorten words by stemming English words. Once the preprocessing has been performed then validaiton has been performed for prediction. Training and testing are the two components that make up validation. Patients are tested on a regular basis, and the classifier is present in the training. Finally resutls are evaluated using performance evaluation metrics (Hussain et al., 2015). Recent works, bias and inconsistency of depression prediction has been reduced by the deep learning technique (Tausczik & Pennebaker, 2010).

In this paper, a Whale Optimization Algorithm based Convolution Neural Networks (WOA-CNN) depression prediction model is introduced to predict depression by employing tweet dataset. Bias, variance, and irreducible errors make up the prediction error. Bias is the difference between the predicted value of the estimate and the target outputs. Variance is the inaccuracy caused by changes in the training set. The irreducible error is made up of unobserved information and observation noise. This system has helped to detect the depressive symptoms of users from Twitter sentiment140 dataset.

Section 1 of the article provides an overview of depression including the causes, symptoms, treatment options, and role of deep learning methods in finding depression from social media. Section 2 describes the recent methods for depression prediction with social media. In section 3 proposed methodology is explained by proposing a novel WOA-CNN optimization technique and other steps for depression prediction. Section 4 provides a comprehensive experimental evaluation, and section 5 concludes with a conclusion and recommendations for further research.

2. Literature Survey

In CLEF in 2018, Wang et al. (2018) proposed a technique for depression and anorexia early risk prediction on social media. Convolutional neural networks (CNNs) and Term Frequency- Inverse Document Frequency (TF-IDF) data are used in the models to find potential patients of two mental illnesses who may have published papers. The models correctly identify anorexia with an ERDE5 of 13.65%, ERDE50 of 11.14%, and F-score of 0.67 and depression with an ERDE5 of 10.81%, ERDE50 of 9.22%, and F-score of 0.37, according to the official assessment.

Sau and Bhakta (2017) constructed an acceptable prediction model based on sociodemographic and health-related factors, to detect anxiety and depression in senior people. Ten classifiers were investigated and evaluated using the ten-fold cross-validation method using a data set of 510 elderly patients. With a prediction accuracy of 89%, the random forest (RF) classifier achieved the best performance. The 110 senior individuals from a second dataset were used to assess the RF model's external validity. It anticipated accuracy to be 91% and had a 10% of false positive (FP) rate when compared to the gold standard instrument.

Machine learning algorithms were used by Priya, Garg, and Tigga (2020) to anticipate stress, despair, and anxiety. To use these techniques, data from employed and unemployed persons from diverse cultures and groups was gathered using the Depression, Anxiety, and Stress Scale (DASS 21) questionnaire. Five different machine learning techniques were used to predict anxiety, depression, and stress on a scale of one to five degrees of severity: Decision Tree (DT), RF, Naive Bayes (NBs), Support Vector Machine (SVM), and K Nearest Neighbor (KNN). These techniques are particularly well suited to psychological issues because of their high accuracy. The confusion matrix showed that classes were out of balance after the application of the different techniques. In order to help choose the optimal accuracy model out of five methods, the F1-score measure was implemented.

Cong et al. (2018) presented an XGBoost-Attention-bidirectional Long Short-term memory (X-A-BiLSTM) Using imbalanced social media data, a deep learning algorithm can identify depression. The two primary components of the X-A-BiLSTM model are an Attention-BiLSTM neural network that enhances classification performance and XGBoost, which is utilised to reduce data imbalance. Reddit Self-Reported Depression Diagnosis (RSDD) database is utilized and included 107,000 matched control users and 9,000 users who claimed to have been diagnosed with depression. On the RSDD database, the technique outperforms earlier modern models by a wide margin.

Orabi et al. (2018) presented a deep neural network architectures to detect mental disorders, specifically depression. With the goal of detecting persons who are depressed from their social media postings, such as tweets, a novel technique to word-embedding optimization for classification is offered. a suggestion for how to better execute two tasks: test generalisation ability on the Bell Lets Talk database and depression detection on the CLPsych2015 database.

Oh et al. (2019) proposed a Deep-Neural-Network (DNN) and Logistic Regression (LR) In 2014, 4949 individuals from the South Korea NHANES (K-NHANES) database and 19,725 participants from the NHANES database's survey data were used to evaluate classifiers.Area under the Receiver Operating Characteristic Curve (AUCs) values of 0.91 and 0.89, respectively, were found by a deep-learning system for diagnosing depression in the NHANES and K-NHANES datasets. The deep learning method was superior than LR in terms of performance (AUC, 0.77). (AUC, 0.74).In both the NHANES and K-NHANES datasets, DNN was able to distinguish depression from other demographic and health characteristics with reasonable accuracy. On a fresh database, across both time and country, the deep-learning system was also able to predict depression rather effectively.

Shah et al. (2020) proposed a hybrid technique to assess textual messages from users to identify depression. For the pilot project, Early Detection of Depression in Conference and Labs of the Evaluation Forum (CLEF) eRisk 2017, deep learning algorithms were built using training data, and their performance was subsequently assessed using test data from the dataset of Reddit. In particular, the Bidirectional Long Short Term Memory (BiLSTM) was proposed, which included a variety of word embedding methods and metadata elements, and it produced favorable results.

Kour and Gupta (2022) Convolutional Neural Network (CNN) and BiLSTM, a proposed combination of two deep learning architectures, perform optimally on a benchmark depression database incorporating tweets and achieve an accuracy of 94.28%.Comparisons are made between the CNN-BiLSTM model, the RNN and CNN models, as well as the standard methods. The proposed approach seems to increase predictive performance, according to experimental findings based on a variety of performance indicators. The stark contrast between the language representation of depressed and non-depressive material was shown to further illuminate the issue using statistical tools and visualization methods.

Baghdadi et al. (2022) Lemmatization, stemming, and several lexical analysis techniques are contrasted in a proposed Arabic twitter preparation system. Twitter data that has been retrieved from the Internet is used in experiments. The data has been annotated by five different people. Using the most recent versions of the Transformers Bidirectional Encoder Representations (BERT) and Universal Sentence Encoder (USE) models, performance measurements are presented on the database. Balanced accuracy, specificity, F1-score, Intersection over Union (IoU), receiver operating characteristic curve (ROC), Youden Index, and weighted sum metrics (WSM) are the performance measures that were assessed. The best WSM for Arabic BERT models is 95.26 %, while the best WSM for USE models is 80.20%. Zogan et al. (2021) proposed a unique computational framework for automatically detecting sadness that first picks pertinent material using a hybrid extractive and abstractive summarizing technique on the order of all user tweets, resulting in a more precise and pertinent content. The material is then transferred to an innovative deep learning framework that combines attention-enhanced Gated Recurrent Units (GRU) models with Convolutional Neural Network (CNN) technology to provide empirical results that are superior to those of strong baselines already in use.

Table 1. **Summary of Literature Review**

Author & Year	Algorithm	Advantage	Disadvantage
Wang et al. (2018)	TF-IDF and CNNs	Early risk detection especially when multiobjectives has been optimized in this work.	Moreover, there is no rejection consensus on the total time of writing.
Sau and Bhakta (2017)	Random Forest (RF)	Inspect an elderly patient for signs of anxiety and sadness.	Doesn't predict anxiety and depression among the younger age patients, and takes more time complete task.
Priya et al. (2020)	machine learning classifiers	High prediction accuracy level.	Imbalance problem in a confusion matrix.
Cong et al. (2018)	X-A-BiLSTM	Behavioral attributes result in better prediction.	Higher computation time to complete task.
Orabi et al. (2018)	CNN& RNN	Proposed system has been worked without any exhaustive feature engineering, and it is being able to update their weights at the time of training.	Data imbalance is more. Detecting depression of Twitter users (i.e., at user level not at post level) by limited data.
Oh et al. (2019)	DNN& Random Forest (RF)	Clearly distinguish depression from other physical and social characteristics.	This system could not evaluate the correlation among the several features and the severity of depression.
Shah et al. (2020)	BiLSTM	Analyze user text messages may help you spot depression.	Too much time passes before their depression is recognised.
Kour and Gupta (2022)	CNN-BiLSTM	The syntactic and semantic content of text data is captured by word embeddings.	The length of a sentence can't be exceeded in one location.
Baghdadi et al. (2022)	BERT& USE	This method can keep track of mental diseases like depression.	Scraped data will not focus in this work.
Zogan et al. (2021)	CNN& GRU	The interplay among automatic summarization, user behavioural representation and model training attains higher depression prediciton. It helps to decrease the curse-of-dimensionality problem.	Summarization takes more time to complete the task.

3. Proposed Methodology

In this paper, initially dataset samples have been collected from Sentiment140. Secondly preprocessing steps like upper case/lower case removal, stemming, stop-words removal, characters removal, and tokenization has been performed to text dataset. Then, feature extraction by Convolution Neural Networks (CNN) has been used to generate word embeddigns feature vector. To help a prediction system perform better, extracted features identify the relevant data dimensions.WOA algorithm optimizes the CNN model for greater accuracy. Finally, the precision, recall, F-measure, and accuracy metrics have been used to evaluate the performance characteristics of the proposed depressive framework. All these steps are illustrated in figure 1.

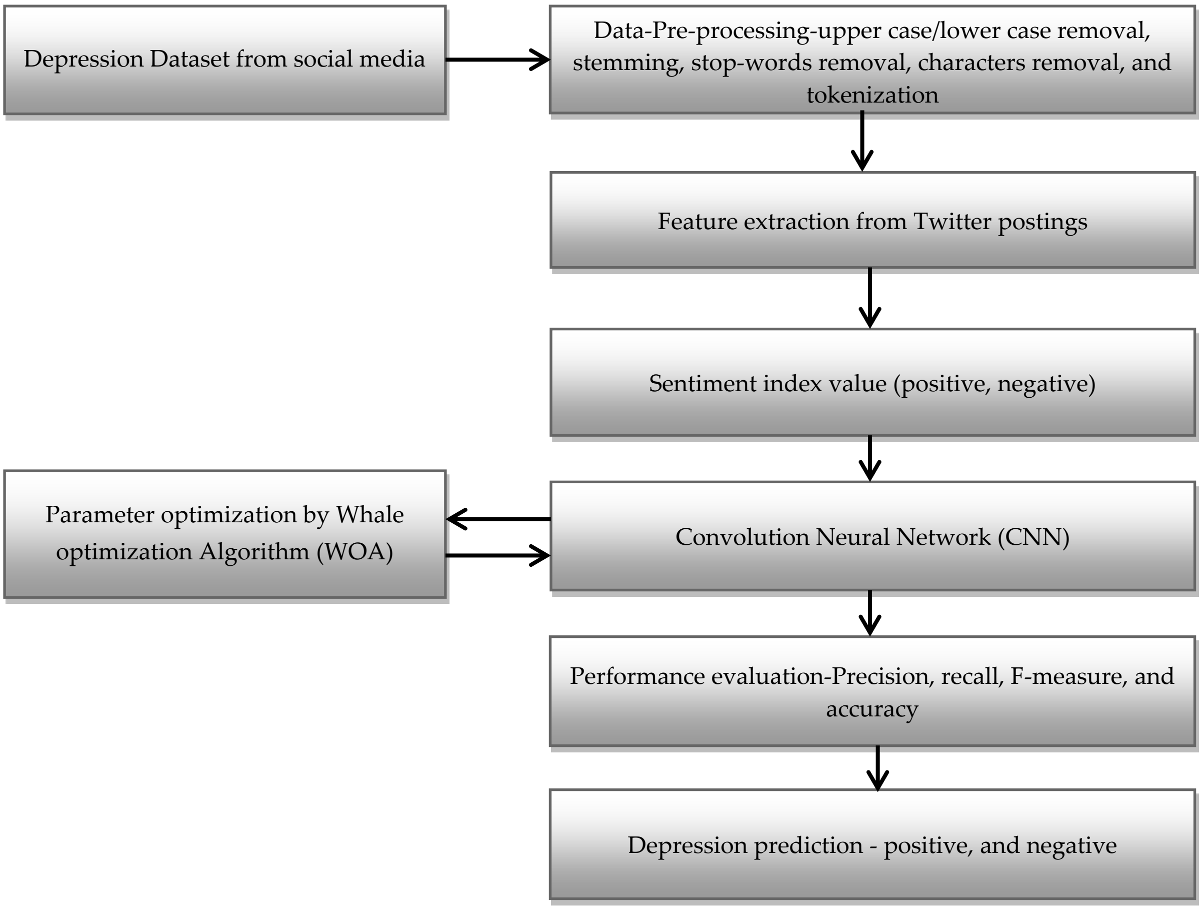

Figure 1. **The Proposed Method's Block Diagram**

3.1. Dataset description

Data may be accessed freely and conveniently on the well-known social media website Twitter. It takes a lot of work to create and validate the phrases that are used as a lexicon for people with mental illness to browse data. The benchmark dataset for depression prediction is collected from https://raw.githubusercontent.com/viritaromero/Detecting-Depression-in-Tweets/master/sentiment_tweets3.csv. A portion of Sentiment140 (8,000 positive tweets) and another database for depressed tweets (2,314 tweets), for a total of 10,314 tweets, were combined to create this database.

3.2. Data pre-processing

As indicated above, the second phase is the collection of data from social media and hence the noise from unstructured data for improving the overall output. The data obtained from the real world are gathered using a variety of techniques, and they are not restricted to a single field; as a consequence, the data are insufficient, illogical, and untrustworthy, and they include mistakes. If such data are directly evaluated, the predictions produced are irrelevant and incorrect. In proposed framework, various steps are used during the pre-processing phase. The following steps are presented for data preprocessing.

Step-1: The texts which are irrelevant to depression have to remove from the tweet text. These irrelevant cases are a form of noise and it has to be removed.The first step gets rid of the user-specified text patterns such as "user handles (@username)", "hashtags (#hashtag)", "URLs", "characters, symbols, and numbers other than alphabets", "empty strings", "drop rows with NaN in the column", "duplicate rows", etc. are to be removed. Every tweet in the dataset is cleaned up, and all URLs are removed. Because URLs are useless for making predictions, they are not considered, and removing them would simplify the computation process. The time, date, numbers, and hashtags should all be removed next. The tweets no longer include the date and time since they are unhelpful for predicting sadness. Similar to how numbers are not a good choice for prediction, hashtags may be utilized, but they are not a good choice either. When a forecast is made using hashtags, it has been shown that accuracy is quite poor.

Step-2: Confusion may arise when some texts are lower and some are in upper cases. So, have to change all the upper case to lower case for better classification.Emojis and superfluous spaces are also removed from the phrase using this technique.

Step-3: The stop-words such as 'a, an, the, into, if, and, as, for' which are also known as conjunctions, prepositions have to be removed. By doing so, the overall dimension of data can be minimized. To exclude stop words from tweet text, the NLTK (Yogish, Manjunath, & Hegadi, 2018) package, which has a collection of stop terms, is used.

Step-4: Once the stop words get removed, whitespaces has been removed.

Step-5: Stemming has to be done which is the procedure of neglecting prefixes and suffixes of particular data. By removing prefixes or suffixes (such as "ize," "ed," "s," "de," etc.) from a word, Porter Stemmer is used to create the root of the word.

Step-6: The cleaned tweets are retrieved and sent to the tokenizer as input in the next phase once all of the tweets have been cleaned.A significant pre-processing step is tokenizing raw text data (Shen et al., 2017). By dividing a longer body of text into smaller lines or words, it uses regular expressions to separate a given string into tokens. Importing the NLTK package is necessary to utilize the various tokenization methods. Providing datasets of cleansed good and negative tweets is the initial step in the tokenizing process.From a collection of texts, it refreshes the internal vocabulary, and based on the frequency of terms, it generates the vocabulary index. As a consequence, the term that appears the most often has the lowest index value. This method thus returns the most words possible together with an index for each word. The texts to sequences() function is the next step of tokenization. It gets information from the stemming, including as many words with an index as possible. Every word in a tweet is to be converted into a series of numbers with the intention of replacing each one with the relevant integer value from the word index dictionary. At this point, the tweets are transformed into integer sequences with different lengths.

3.3. Feature extraction

The quality of feature extraction is significantly improved by processing data in accordance with the needs. To extract significant and pertinent characteristics, feature extraction is used on the pre-processed data.To improve classification performance; extracted features identify the relevant data dimensions. A word that is accessible in very sparse vectors may be projected down into a low dimensional embedding vector using the technique of embedding, which has been developed to cope with huge data. Dense or small-scale examples of high-dimensional input vectors are called embeddings (Alvi, Talukder, & Uddin, 2022; Fatima et al., 2019). Word embeddings (Mandelbaum & Shalev, 2016) are taught to place similar-meaning words next to one another and produce vectors with almost identical representations. For instance, the keywords "joyful" and "miserable" have quite different semantic connotations. In order to create more distinguishable characteristics out of the tokenized numerical vectors, they will be represented widely away from one another in the geometric space. In order to capture the semantic link between the corresponding word vectors, these vectors are modified with the aid of an embedded layer. When compared to embedding vectors in the embedding space, which learn relationships between words based on the distance between two vectors, the initial tokenized vector lacks relationships between various words. More separate features are retrieved with each iteration of the training, increasing the CNN's capacity for prediction.The numerical vectors for each preprocessed data point are calculated using the word embeddings approach. In order to create word indexes, first turned all of the sample text's words into sequences. Through the use of the Keras text tokenizer, these indexes are obtained. Tokenizer will never provide a zero index value to a word, and the length of the vocabulary will be adjusted to reflect this. After that, each every word in the database is given its own distinct index, and this index is then used in the formation of numeric vectors for each and every text sample.

3.4. Sentiment index value

Initially, the words are set as positive and negative cases. The positive is marked based on calculating the positive terms of lexicon words which must be greater than negative terms and vice versa in the case of negative words. It is based on the lexicon dictionary the localized weight is calculated for positive, and negative. The table above displays a list of sentimental terms that can be used to identify depression on social media. User-level sentiment index can be calibrated using the equation(1).

$$\mathrm{Ci, j} = \mathrm{a(i, j)} + \mathrm{b(i, j)} \tag{1}$$

where an (i,j) is the localized weight factor for positive words, b(i,j) is the localized weight factor for negative words. The indication of the inverse documentary (bi) for negative words is indicated by equation (2),

$$bi = \{ {}^{1+\log(ni) if ni>1}_{0} \quad (2)$$

where ni is the total number of words. Now, the globalized weight factor (zi) with the indication of an inverse document is indicated byequation (3),

$$zi = \log \frac{N}{ui} \quad (3)$$

wherein N is total texts, ui is total users.

$$ui = \frac{1}{\sqrt{\sum_{i}^{N} (aiz_i)^{2^n}}} \quad (4)$$

Once, the weight-based sentiment index is calibrated, the user data is filtered and arranged to make it ready for medical help.

3.5. Depression prediction

Whale Optimization Algorithm based Convolution Neural Networks (WOA-CNN) depression prediction involves the classification of text by categorizing it into positive, and negative. The classification of text is as follows.

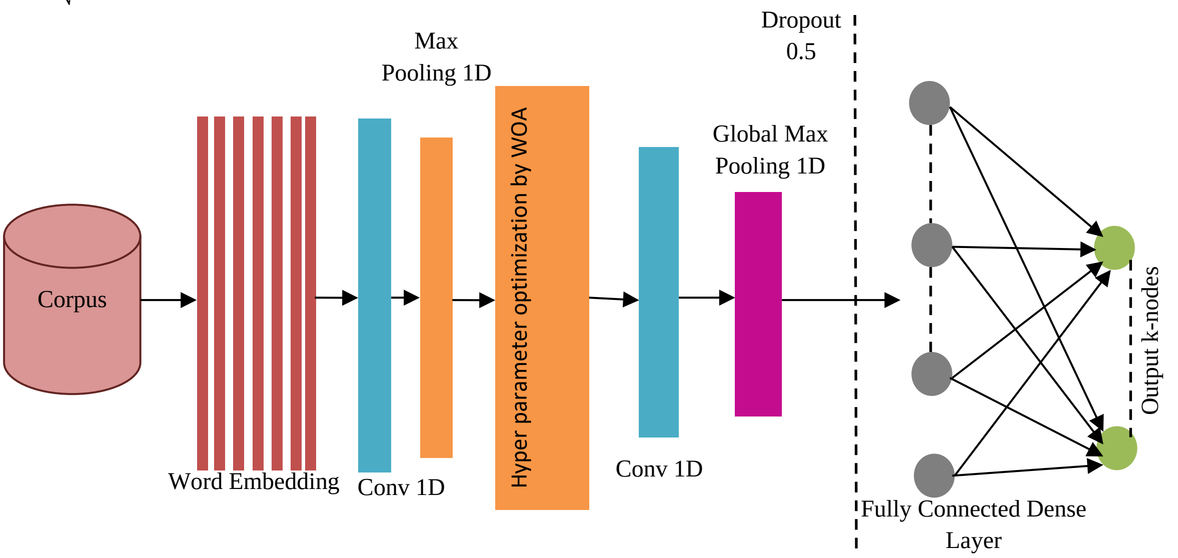

Figure 2. **The Proposed WOA-CNN Model Architecture**

The working of WOA-CNN architecture is as follows. Create a CNN layer first to extract words that describe sadness from user postings on social media. As the first layer, the embedding layer transforms the inputted text into a word vector, which is designated as $W \in R^{k \times m}$, where the text's word count cap is k, and word embedding's dimension is m. The maximum sequence length is set to be equal to the longest possible length of text in the database, and the embedding dimension is set at 300.

$$W_k = v_1^w \oplus v_2^w \oplus \ldots \oplus v_r^w \quad (5)$$

In equation (5), $W_k$the words and are incorporated into $\oplus$is the operation of concatenation. Wkbuilt by combining togetherv_r^w where $v_r^w \in R^m$is the rth word in the depressed text's m-dimensional text vector.

Extraction of multi-resolution features from the input matrix is the fundamental function of the convolution layer. To get different kinds of characteristics, several kinds of filters are used. The purpose of the convolutional layer is to identify patterns, also known as discriminative word sequences, that are present in the input word-vectors and that are consistent throughout the training text for depression.Next, inputs are convolved using a filter matrix to create a new feature mapw $\in \mathbb{R}^{h \times m}$, where the convolution's size is *h*. Equation (6) describes the computation of the convolution,

$$c_i = f\left(\sum_{j,k} w_{j,k} \left(W_{[i:i+h-1]}\right)_{j,k} we_{j,k} + b\right) \quad (6)$$

Where $b \in \mathbb{R}^h$a biased term, $we_{j,k} \in \mathbb{R}^{h \times m}$is the kernel's weight, and$f(x)$ is an activation function that is nonlinear. The output is the product of an input matrix W and a filter matrix w, element by element, This may then be combined into one value and shown as a feature map $c = [c_1, c_2, \ldots, c_P]$, where $c \in \mathbb{R}^{n-h+1}$. The model may utilize many filters of varying lengths that run in parallel to generate rich feature representation.

The output of each CNN neuron is subjected to the ReLu activation function. For the network to remain nonlinear, this function changes all negative values to zero. The output shape of the CNN layer after ReLu activation is identical to the input layer's geometry. By merging the scores from each filter, a pooling layer aims to further abstract the characteristics. Each feature map in this model has two max-pooling layers applied to it. By choosing the greatest value on each vector dimension, this layer highlights the key characteristic. The input characteristics' size therefore shrinks significantly. The features will be reduced by kernel/pool size (p), where p equals 2, in the layer's output. Kernel = 5 with "global max pooling" of various sizes. Dropping the input features with values smaller than the dropout rate using the dropout layer helps to minimize overfitting. The dropout rate d for CNN model is 0.5. A dense layer is the final layer of this model which gives final result. It is followed by a softmax activation function has been used for depression prediction.

In this work, the hyper parameters of CNN classifier such as kernel/pool size (p), Number of filters, size of kernel, weight, and bias has been optimized using WOA. To reduce the error of classifier, the CNN parameters are optimized using the whale optimization (Aljarah, Faris, & Mirjalili, 2018; Gharehchopogh & Gholizadeh, 2019). It is based on the activity of whales with humpbacks. These whales start to search for their prey, encircle the prey, and bubble net the prey for tuning parameters of CNN classifier for depression prediction.

The steps of this algorithm for extracting depression words from social media are as follows,

Step-1: Initialization: Initially, the randomized weight values are generated with resultant spaces. They are indicated by w_i, wherein "n" denotes the number of, i=1, 2, 3, and so on weights with initialization factors A, B, and C.

Step-2: Fitness calculation: This calculation is done for 100 groups of depression words using the below indicated equation with the finest degree point exposure.

$$F_i=\min (MSE) \quad (7)$$

Minimum Square Error (MSE) is done by finding the distance between two words and hence it is done for 100 groups of words. Provided a string s and two words w_1 and w_2 existent in S, the distance computation is as follows. The goal is to calculate the shortest distance between w_1 and w_2. The distance between first and second words is measured in steps or words. Make a string with Hamming Distance equal to half the distance among words w_1 and w_2.The character with the shortest distance to every other character is noted.

Step-3: Updating: After finding the fitness for every 100 groups of words, the updating process has to be done to determine the overall weight.

Step-4 Encircling the prey: From the numerous text (or) numerous tweets on Twitter, the selection has to be done by proper documentation of humpback whales. This is followed by updating the operator at each location using the below equation (8),

$$UpD=A.Best\ w\ (t)-d(t) \quad (8)$$

$$D(t+1)=d(best)\ (t)-A\ UpD \quad (9)$$

where each word's location is given by d, its vector by w, and t is the reputation state.The convolutional vector is calibrated as follows:

$$U'=2u.r'-b \quad (10)$$

$$V'=2.r' \quad (11)$$

Where u is the repeated sequence and b is the linear product.

Step-5 Shrinking method: For undergoing this method the entire group of depression words has to shrink with the differential rate of γ. The interval for γ can be set as [-α,α].

Step-6 Spiral method: This is done by making the helix based graph with updated process of w(t+1), w(t), w(t-1) factors.

Step-7 Searching the prey: once the encircling is done, the prey searching process restarts with the coordination of location factor with manipulation range. As a result, the searching is done, and the globalized search factor has to be found with the notation of the numerical factor.

Step 8- Termination criteria: When the best weight point is found to satisfy a closure principle, the WOA algorithm is finished. The simulation parameters of WOA are discussed in table 1.

Table 1. **Simulation Parameters of WOA**

Parameter	Description
N_{pop}= Population size	300
T_{max}= generations Number	140
P_{se}= Shrinking encircling operation probability	0.04
P_{su}= Spiral updating operation probability	0.05
P_{rs}= Random searching operation probability	0.3

Algorithm 1. **Whale Optimization Algorithm (WOA)**

```
Initialization of parameters
p←positive, n←negative,LD←Lexicon Dictionary,
D←database, T←word,
Process the data
Classify (data)
p=0, n=0
For
    Word(T) in database (D)
If
    LD=p then
    p=p+1
Else
    LD=n then
    n=n+1
end if
end for
updating←UpD=A.Best w (t)-d(t)
Fitness calculation
Fi=min (MSE)
if p>n then
    return "positive words"
else if p<n then
    return "negative words"
end if
end process
```

4. Experimentation results

The dataset is obtained from the https://raw.githubusercontent.com/viritaromero/Detecting-Depression-in-Tweets/master/sentiment_tweets3.csv. It has been also used for this dataset collection. A part of the Sentiment140 (8,000 positive tweets) and another database for depressed tweets (2,314 tweets) were combined to produce this database, which has a total of 10,314 tweets. Online depression data undergo the following process of preprocessing, feature extraction, and sentiment classification. A comparison of proposed WOA-CNN method is done with the existing algorithms like BiLSTM, X-A-BiLSTM, andCNN-BiLSTM concerning many parametric measures such as precision, recall, F-measure, Accuracy, and convergence speed. To approve the proposed Whale Optimization algorithm based on Convolution Neural Networks (WO-CNN) is better than the present techniques, graphs are provided.

Precision is often referred to as a positive predictive value, and the computation is shown in equation (12),

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (12)$$

Recall is also known as sensitivity, which refers to the total number of relevant occurrences that were retrieved as indicated in the equation (13),

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (13)$$

According to the equation (13), the F-measure is the harmonic mean of accuracy and recall,

$$\text{F} - \text{measure} = \frac{2*\text{precision}*\text{recall}}{\text{Precision+Recall}} \quad (13)$$

Accuracy is seen as a useful measure for evaluating performance. For data processing, accuracy displays accurately classified data and false positive data. It is indicated in the equation (14)

$$\text{Accuracy} = \frac{\text{TP+TN}}{(\text{TP+TN+FP+FN})} \quad (14)$$

TP is used to indicate true positive values,

In the equation (14) above, the letters TP, TN, FP, FP, and FN stand for true positive, true negative, false positive, and false negative values, respectively.

Table 2. **Evaluation Analysis of Deep Learning Methods Vs. Number of Iterations**

Number of iterations	Precision (%)			
	BiLSTM	X-A-BiLSTM	CNN-BiLSTM	WOA-CNN
100	72.58	75.78	78.17	84.67
200	74.81	77.19	82.70	85.63
300	77.41	82.67	84.93	88.47
400	79.50	84.28	87.74	89.63
500	82.12	86.65	88.40	90.76
Number of iterations	**Recall (%)**			
	BiLSTM	X-A-BiLSTM	CNN-BiLSTM	WOA-CNN
100	75.17	77.11	80.35	83.79
200	77.23	79.65	83.18	86.36
300	79.12	82.61	85.63	88.42
400	82.71	84.50	87.49	90.66
500	84.80	86.18	90.13	92.89
Number of iterations	**F-Measure (%)**			
	BiLSTM	X-A-BiLSTM	CNN-BiLSTM	WOA-CNN
100	73.87	76.44	79.26	84.23
200	76.02	78.42	82.94	85.99
300	78.26	82.64	85.28	88.44
400	81.10	84.39	87.61	90.14
500	83.46	86.41	89.26	91.82
Number of iterations	**Accuracy (%)**			
	BiLSTM	X-A-BiLSTM	CNN-BiLSTM	WOA-CNN
100	76.61	80.55	82.82	88.45
200	78.63	82.12	84.51	90.69
300	81.65	84.11	86.71	92.34
400	83.71	85.45	88.92	92.35
500	85.20	87.44	90.21	93.03

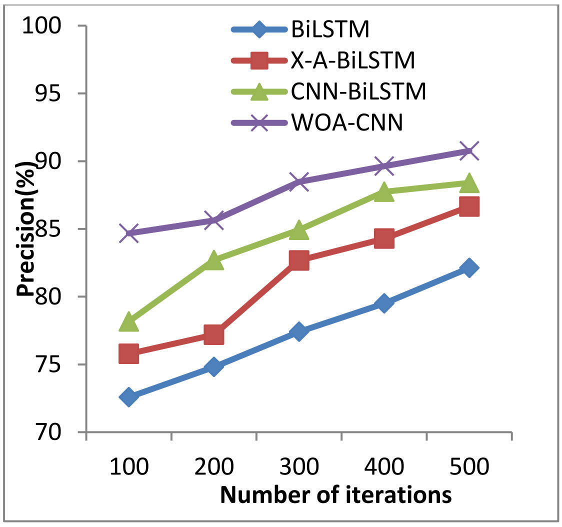

Figure 3. **Precision Results Comparison of Classifiers**

Figure 3 shows the performance comparison of precision with respect to classifiers such as BiLSTM, X-A-BiLSTM, CNN- BiLSTM, and proposed WOA-CNN classifier by number of iterations. The proposed classifier has higher precision results of 90.76%, whereas other methods have 82.12%, 86.65%, and 88.40% for BiLSTM, X-A-BiLSTM, and CNN-BiLSTM (See Table 2) for 500 no.of iterations.

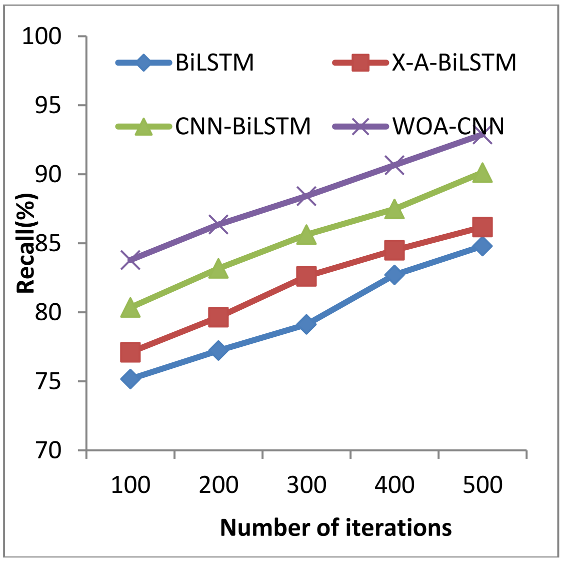

Figure 4. **Comparison of Classifier Recall Results**

Figure 4 shows the recall results comparison of classifiers such as BiLSTM, X-A-BiLSTM, CNN-BiLSTM, and proposed WOA-CNN classifier by number of iterations. The proposed classifier has higher recall results of 92.89%, whereas other methods have 84.80%, 86.18%, and 90.13% for BiLSTM, X-A-BiLSTM, and CNN- BiLSTM (See Table 2) for 500 no.of iterations.

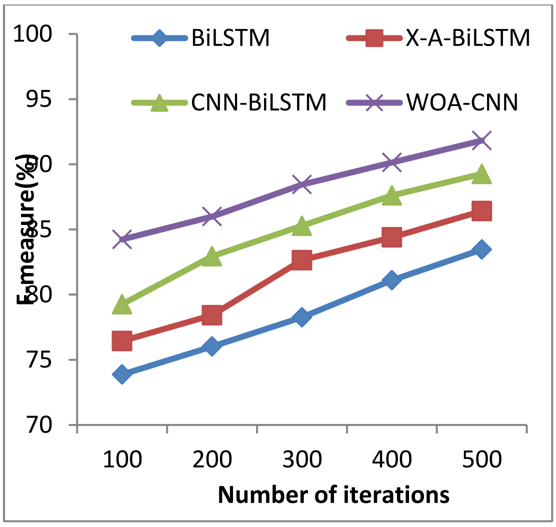

Figure 5. **F-Measure Results Comparison of Classifiers**

Figure 5 compares the results of classifiers like BiLSTM using the F-measure, X-A-BiLSTM, CNN-BiLSTM, and proposed WOA-CNN classifier by number of iterations. The proposed classifier has higher F-measure results of 91.82%, whereas other methods have 83.46%, 86.41%, and 89.26% for BiLSTM, X-A-BiLSTM, and CNN- BiLSTM (See Table 2) for 500 no.of iterations.

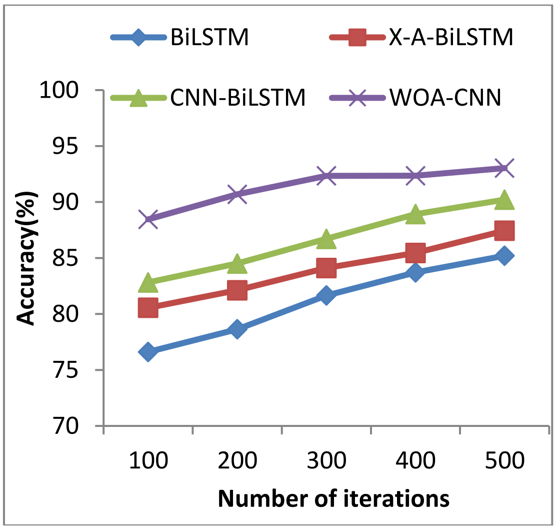

Figure 6. **A Comparison of Classifiers' Accuracy Results**

The comparison of accuracy results is shown in Figure 6 classifiers such as BiLSTM, X-A-BiLSTM, CNN- BiLSTM, and proposed WOA-CNN classifier by number of iterations. The proposed classifier has higher accuracy results of 93.03%, whereas other methods have 85.20%, 87.44%, and 90.21% for BiLSTM, X-A-BiLSTM, and CNN- BiLSTM (See Table 2) for 500 no.of iterations.

5. Conclusion and future work

This research proposes a strategy for detecting depressed people in social networks using sentiment analysis and data mining. By using features from the depression detection model, the sentiment analysis approach pays special attention to the characteristics of depression. The Whale Optimization Algorithm based Convolution Neural Networks (WOA-CNN) method is introduced in this study as a unique depression detection technique. Shrinking encircling, spiral updating, and random searching are three optimization techniques used in WOA-CNN to replicate humpback whale hunting behavior and identify communities in networks. This has been used for the optimization of parameters in CNN classifier. As a result, the accuracy of the proposed WOA-CNN algorithm achieves higher precision, recall, f-measure, and accuracy of about 90.76%, 92.89%, 91.82%, and 93.03%, for about 500 no. of iteration. Studies on artificial and real-world networks show that proposed system is capable of discovery. Nevertheless, the effectiveness of WOA-CNN improves the accuracy, but it takes a long time in the face of a huge search space. WOA-CNN framework has focus on all tweets information during training which significantly leads to lesser prediction performance due the curse-of-dimensionality problem. This will be kept as scope of future work for reducing the dimensionality of features from the tweet dataset. Evolutionary algorithms has been introduced for solving dimensionality problem to increase the depression prediction results, and to minimize the running time of the overall system.

References

[1] Aljarah, I., Faris, H., & Mirjalili, S. (2018). Optimizing connection weights in neural networks using the whale optimization algorithm. *Soft Computing, 22*(1), 1-15. https://doi.org/10.1007/s00500-016-2442-1

[2] Alvi, N., Talukder, K. H., & Uddin, A. H. (2022). Sentiment Analysis of Bangla Text Using Gated Recurrent Neural Network. In *International Conference on Innovative Computing and Communications* (pp. 77-86). Springer. https://doi.org/10.1007/978-981-16-2597-8_7

[3] Baghdadi, N. A., Malki, A., Balaha, H. M., AbdulAzeem, Y., Badawy, M., & Elhosseini, M. (2022). An optimized deep learning approach for suicide detection through Arabic tweets. *PeerJ Computer Science, 8*, e1070. https://doi.org/10.7717/peerj-cs.1070

[4] Cong, Q., Feng, Z., Li, F., Xiang, Y., Rao, G., & Tao, C. (2018). XA-BiLSTM: A deep learning approach for depression detection in imbalanced data. In *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 1624-1627). IEEE. https://doi.org/10.1109/BIBM.2018.8621230

[5] Fatima, I., Abbasi, B. U. D., Khan, S., Al-Saeed, M., Ahmad, H. F., & Mumtaz, R. (2019). Prediction of postpartum depression using machine learning techniques from social media text. *Expert Systems, 36*(4), e12409. https://doi.org/10.1111/exsy.12409

[6] Fleuren, W. W., & Alkema, W. (2015). Application of text mining in the biomedical domain. *Methods, 74*, 97-106. https://doi.org/10.1016/j.ymeth.2015.01.015

[7] Gharehchopogh, F. S., & Gholizadeh, H. (2019). A comprehensive survey: Whale Optimization Algorithm and its applications. *Swarm and Evolutionary Computation, 48*, 1-24. https://doi.org/10.1016/j.swevo.2019.03.004

[8] Hussain, J., Ali, M., Bilal, H. S. M., Afzal, M., Ahmad, H. F., Banos, O., & Lee, S. (2015). SNS based predictive model for depression. In *International Conference on Smart Homes and Health Telematics* (pp. 349-354). Springer. https://doi.org/10.1007/978-3-319-19312-0_34

[9] Kour, H., & Gupta, M. K. (2022). An hybrid deep learning approach for depression prediction from user tweets using feature-rich CNN and bi-directional LSTM. *Multimedia Tools and Applications, 81*, 23649–23685. https://doi.org/10.1007/s11042-022-12648-y

[10] Kumar, B. S., & Ravi, V. (2016). A survey of the applications of text mining in financial domain. *Knowledge-Based Systems, 114*, 128-147. https://doi.org/10.1016/j.knosys.2016.10.003

[11] Lim, S., Tucker, C. S., & Kumara, S. (2017). An unsupervised machine learning model for discovering latent infectious diseases using social media data. *Journal of biomedical informatics, 66*, 82-94. https://doi.org/10.1016/j.jbi.2016.12.007

[12] Mandelbaum, A., & Shalev, A. (2016). Word embeddings and their use in sentence classification tasks. *arXiv preprint arXiv:1610.08229*, 1-16. https://doi.org/10.48550/arXiv.1610.08229

[13] Oh, J., Yun, K., Maoz, U., Kim, T.-S., & Chae, J.-H. (2019). Identifying depression in the National Health and Nutrition Examination Survey data using a deep learning algorithm. *Journal of affective disorders, 257*, 623-631. https://doi.org/10.1016/j.jad.2019.06.034

[14] Orabi, A. H., Buddhitha, P., Orabi, M. H., & Inkpen, D. (2018). Deep learning for depression detection of twitter users. In *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic* (pp. 88-97). Association for Computational Linguistics. http://dx.doi.org/10.18653/v1/W18-0609

[15] Priya, A., Garg, S., & Tigga, N. P. (2020). Predicting anxiety, depression and stress in modern life using machine learning algorithms. *Procedia Computer Science, 167*, 1258-1267. https://doi.org/10.1016/j.procs.2020.03.442

[16] Sánchez, D., Batet, M., Martínez, S., & Domingo-Ferrer, J. (2015). Semantic variance: an intuitive measure for ontology accuracy evaluation. *Engineering Applications of Artificial Intelligence, 39*, 89-99. https://doi.org/10.1016/j.engappai.2014.11.012

[17] Sau, A., & Bhakta, I. (2017). Predicting anxiety and depression in elderly patients using machine learning technology. *Healthcare Technology Letters, 4*(6), 238-243. https://doi.org/10.1049/htl.2016.0096

[18] Shah, F. M., Ahmed, F., Joy, S. K. S., Ahmed, S., Sadek, S., Shil, R., & Kabir, M. H. (2020). Early depression detection from social network using deep learning techniques. In *2020 IEEE Region 10 Symposium (TENSYMP)* (pp. 823-826). IEEE. https://doi.org/10.1109/TENSYMP50017.2020.9231008

[19] Shen, G., Jia, J., Nie, L., Feng, F., Zhang, C., Hu, T., Chua, T.-S., & Zhu, W. (2017). Depression detection via harvesting social media: A multimodal dictionary learning solution. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence* (pp. 3838-3844). IJCAI. http://hcsi.cs.tsinghua.edu.cn/Paper/Paper17/IJCAI17-SHENGUANGYAO.pdf

[20] Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology, 29*(1), 24-54. https://doi.org/10.1177/0261927X09351676

[21] Wakefield, J. C., Schmitz, M. F., & Baer, J. C. (2010). Does the DSM-IV clinical significance criterion for major depression reduce false positives? Evidence from the National Comorbidity Survey Replication. *American Journal of Psychiatry, 167*(3), 298-304. https://doi.org/10.1176/appi.ajp.2009.09040553

[22] Wang, Y.-T., Huang, H.-H., Chen, H.-H., & Chen, H. (2018). A Neural Network Approach to Early Risk Detection of Depression and Anorexia on Social Media Text. *CLEF (Working Notes)*, 1-8. http://nlg.csie.ntu.edu.tw/~hhhuang/docs/clef2018a.pdf

[23] Yadav, P. (2014). Mutual information-aided Semantic Approach for Medical ontology construction. *Internatinal Journal Publications of Problems and Applications in Engineering Research, 5*, 322-326. http://ijpaper.com/wp-content/uploads/2012/08/IJPAPER-Y14-12069.pdf

[24] Yogish, D., Manjunath, T., & Hegadi, R. S. (2018). Review on natural language processing trends and techniques using NLTK. In *International Conference on Recent Trends in Image Processing and Pattern Recognition* (pp. 589-606). Springer. https://doi.org/10.1007/978-981-13-9187-3_53

[25] Zogan, H., Razzak, I., Jameel, S., & Xu, G. (2021). Depressionnet: learning multi-modalities with user post summarization for depression detection on social media. In *proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval* (pp. 133-142). ACM Digital Library. https://doi.org/10.1145/3404835.3462938