Ensemble Neural Models for Depressive Tendency Prediction Based on Social Media Activity of Twitter Users

Gurdeep Saini, Naveen Yadav, and Sowmya Kamath S

1 Introduction

Depression is a mental affliction that is often overlooked and is typically regarded as a non-serious condition in most countries. However, chronic depression causes the manifestation of various physical symptoms such as loss of appetite, reduction in attention and concentration, lack of self-confidence, and disturbed sleep. Many life experiences can trigger a bout of depression, such as losing a job, losing a loved one, family problems, health issues, and other tough situations such as COVID-19 outbreak. The pandemic has caused several punitive actions like the nationwide lockdown and physical distancing due to which many people have been leading isolated lives different from their normal life, making them feel lonely, emotionally distanced, and overwhelmed.

The World Health Organization (WHO) reports that 264 million people worldwide suffer from depression., while 1 in 13 globally suffers from anxiety [1]. Depression is often not diagnosed and is often attributed to other causes like fatigue or tiredness. Depression, at its worst, may lead to suicide, with the World Health Organization [2] reporting that approximately 8,00,000 depressed individuals commit suicide each year. Researchers have attempted to gather information on the problem through survey-based approaches using online question-answer and phone calls [3] However, the major limitations of these approaches are that the collected information is primarily incomplete dependent on the sampling processes used and due to the introduction of sampling bias.

In recent years, research studies have attempted to study mental health and mood using social media data [4–8]. Some approaches are based on supervised techniques [7, 9–12], while others have used lexicon-based methods [13, 14]. In supervised

G. Saini (✉) · N. Yadav · S. Kamath S
Department of Information Technology, Healthcare Analytics and Language Engineering (HALE), National Institute of Technology Karnataka, Surathkal, Mangalore 575025, India
e-mail: sowmyakamath@nitk.edu.in

learning-based approaches, machine learning algorithms such as support vector machine and decision tree have been used for sentiment analysis [15, 16], event analysis [17, 18], personalized recommendation [16], etc., with good results. In lexicon-based methods, researchers used the most important words, i.e., the words which have high frequencies compared to the rest of the words, for classifying the testing tweets.

These works, however, are limited due to a lack of consideration to an individual's forms of expression, followers and influencing network, user involvement, and emotion [15]. These aspects can serve as discriminating features to recognize depression-indicative posts very effectively. Social media networks like Twitter and Facebook have become very good outlets for sharing updates on one's life. Such social behavior on short messaging platforms like Twitter can be an excellent resource that can be used to generate insights into user behaviors, emotions, and feelings, that may be indicative of their mental health. For instance, consider the following news headlines – "*Twitter fail: Teen Sent 144 Tweets Before Committing Suicide & No One Helped*" and "*Jim Carrey's Girlfriend: Her Last Tweet Before Committing Suicide 'Signing Off'*". These underscore the long-term depressive tendencies of the individual to a large audience on the respective social networks, which could have been used to prevent their eventual suicides. Several such works have shown that social media analysis is a good way to detect and address depressive tendencies. Challenges like low accuracy and labor-intensive methodology still persist, and in our work, we aim to address these limitations.

Our work focuses on identifying depressive tendencies based on users social media activity. To develop the proposed model, we utilized a large, public dataset called Kaggle Sentiment140, which contains 1,600,000 tweets with labels negative, neutral, and positive. First, we experimented with supervised learning models such as Support Vector Machine, Random Forests, and XGB for the task, while incorporating Term weight modeling techniques to boost prediction accuracy. We also present an ensemble model built on Convolutional Neural Networks and LSTM for further improving the accuracy of depression prediction. The rest of this article is organized as follows. Section 2 discusses several state-of-the-art methods for the tasks of depression detection. The details of the proposed machine learning models and deep learning model for the task of depression detection are discussed in detail in Sect. 3. The experimental evaluation and observed outcome are discussed in Sect. 4, later conclusion and future work is discussed.

2 Related Work

Studying social behavior for analyzing and understanding underlying patterns for enabling tasks like opinion mining, purchasing habits, bias estimation, and so on has become an increasingly popular research area. Several works exist that attempt to utilize social communication for identifying disease parameters for chronic illnesses like depression [19–22]. Choudhury et al. [9] state that successfully combating depres-

sion is a true measure of an individual's and society's well-being. A large number of people suffer from the negative symptoms of depression, but only a small percentage are diagnosed and receive adequate care. Researchers have explored social media platforms such as Twitter, Facebook, etc., and users' internet-based interactions to detect and assess the symptoms of serious depression in people. Through the use of text used by users in their posts, latent patterns associated with users' mindset, social contacts, emotional framework, self-esteem, hatred, etc., have been analyzed and studied.

Choudhury et al. [23] recognized online networking as a viable public-health tool, concentrating on the use of Twitter to develop statistical models for the impact of delivery on new mothers' behavior and temperament. They used Twitter posts to monitor 376 mothers' postpartum shifts in terms of social interaction, feeling, casual culture, and phonetic style. O'Dea et al. [24] found that Twitter is being increasingly studied as a means of detecting psychological well-being, such as stress and suicidal tendencies, in the general population. Using both human coders and a programmed computer classifier, They observed that the level of anxiety shown in suicide-related tweets may be detected accurately. Zhang et al. [25] showed that if people at high risk of suicide can be identified by online networking such as microblogging, a complex intervention mechanism to save their lives can be implemented. They used a dataset consisting of posts from 1041 Weibo users and NLP techniques like Latent Dirichlet Allocation (LDA) and Linguistic Inquiry Word Count (LIWC) were used to extract linguistic features. Their experiments showed that LDA outperformed all other models in the task of finding the topic that relates to suicide probability.

Paul and Dredze [26] developed a prediction model for modeling the occurrence of various diseases in the public based on the wordings used and emotion expressed in twitter posts. Sadilek and Kautz [27] proved the correlation between social media data and diagnostic influenza patient. Experiments performed on 2.5 million geo-tagged twitter messages, using Sadilek et al., showed interesting insights into the spread of influenza. They discovered that as the number of sick friends increases, the probability of falling ill increases exponentially. Billing and Moos [28] proved that users' stress levels can be predicted based on their twitter feed using various methods like linguistic analysis and emotion analysis.

Aldarwish and Ahmed [29] proposed a depression detection technique using Rapid-Miner. They used two classifier Naive Bayes classifier and SVM classifier. on two data sets containing 2073 depressed posts and social media posts. The limitations of the approach are that the models were manually trained, making it difficult to train and integrate into the model. Husain [30] used user generated content from Facebook, then they labeled the words from the post in the training phase. For testing, text classification algorithms are applied to detect the positive and negative class using SVM classifier. Biradar and Totad [31] proposed Sentistrength sentiment analysis and back propagation neural network for depression prediction. The advantage of using back propagation neural network is that it is fast, simple, and no other parameter is required except inputs. However, it cannot learn non-linearly tasks, which is a significant limitation of its approach.

3 Proposed Methodology

Different mechanisms outlined as part of the proposed strategy for depression prediction are discussed in this section.

3.1 Data Preprocessing

The Sentiment140 dataset containing 1,600,000 tweets was used for the evaluation. The dataset contains a target field—if the target is 0, it means the tweet is negative, if target is 2, then it means the tweet is neutral, and if the target is 4, then the tweet is positive. During data preprocessing, all URLs, unnecessary articles, punctuation, alphanumeric characters, and stopwords except first, second and third pronouns were removed. The NLTK tweet tokenizer was used to tokenize the messages. After tokenizing, we utilized the training dataset to create a final vocabulary which contains 242,660 unique tokens and is used to encode the sentences as series of indices.

3.2 Dealing with Multi-class Data

The dataset contains multiple classes, in order to categorize the preprocessed documents in the relevant classes, we adopt hyperplane- based concepts. The vocabulary is represented in an n-dimensional hyperspace using which classification can be performed. In the high-dimensional space, Support Vector Machine (SVM) is used to train and test the model. As it is a supervised learning model that maps two separate classes in a high-dimensional space, SVM has the potential to reduce overfitting difficulties to some level. To avoid the risk of overfitting [32], it may modify many features while maintaining great performance.

Let us consider datapoints $X_1, X_2, X_2 \ldots X_n$ that are to be mapped to a n-dimensional hyperspace. Equation (3) represents the hyperplane in n dimensions.

$$\alpha_0 + \alpha_1 * X_1 + \alpha_2 * X_2 + \alpha_3 * X_3 + \alpha_4 * X_4 + \cdots + \alpha_n * X_n = 0 \qquad (1)$$

where, α_0 is intercept, α_1 is coefficient of first axis and α_1 is coefficient of nth axis. To adjust this $n - 1$ dimensional plane in n dimensions, we need to adjust the weight such that for any point which belongs to output class, the value obtained with Eq. (2) should be less than 0. If the point (e.g. here point has n values in the n dimensional space, Tweet1 vector and Tweet2 vector are two points) lies in positive class, then Eq. (2) will give a value greater than or equal zero.

$$Weight = \alpha_0 + \alpha_1 * X_1 + \alpha_2 * X_2 + \alpha_3 * X_3 + \alpha_4 * X_4 + \cdots + \alpha_n * X_n \qquad (2)$$

Thus, we can conclude that if some feature maps to a negative class, its corresponding α value will be negative, and similarly for the feature which is inclined to positive class has corresponding α value as positive. During training, the SVM classifier adjusts the α values by using the maximal margin from the vectors, as per Eqs. (4) and (5), where M is the maximum achievable margin form the hyperplane to any of its side and m is the total training vectors, $Y = \{-1, 1\}$ depend on the class of vector, -1 for negative class and 1 for positive class, these two ensure that maximum margin is maintained between two classes.

$$M = \alpha_0 + \alpha_1 + \alpha_2 + \alpha_3 + \cdots + \alpha_n \tag{3}$$

$$Y_i(\alpha_0 + \alpha_1 * X_{i1} + \alpha_2 * X_{i2} + \alpha_3 * X_{i3} + \cdots + \alpha_n * X_{in} >= M \tag{4}$$

In the testing phase, we calculate the value of Eq. (5) for the testing vector. The vector belongs to the negative class if $F(X)$ is negative, and the testing vector belongs to the positive class if $F(X)$ is positive.

$$F(X) = \alpha_0 + \alpha_1 * X_1 + \alpha_2 * X_2 + \alpha_3 * X_3 + \cdots + \alpha_n * X_n \tag{5}$$

The process is performed as follows—firstly, we take the label data set which contain both positive and negative class, and split the dataset such that 80% is used for training and 20% is used for testing. After preprocessing, the data is fed into the SVM classifier, Random forest classifier and XGB Classifier for training/testing the model. The metric Social Media Depression Index (SMBI) is generated to measure the depression level of users in order to observe the performance of the trained ML models. The SMDI value is calculated for both men and women separately and comparative analysis of SMDI is undertaken in the night and day window. Performance obtained with ML models was lower in accuracy due to which we also experimented with deep neural models such as Convolutional Neural Networks (CNN) and Long Short-term Memory (LSTM).

4 Implementation Specifics

4.1 Depression Behavior Identification

For this task, we firstly identify specific keywords that are typically indicative of depressions, i.e., words that indicate unhappiness, negative or abnormal emotions in the post. Some such words are shown in Table 1. Time is also important factor to determine whether a user is in depression or not. To capture this aspect, we used a day and night window, to also model the fact that people are typically more depressed at night when compared to daytime due to factors like loneliness, darkness, hopelessness, isolation, fatigue, etc.

Table 1 Depression-indicative terms

Class	Type	Sample terms
1	Depression-indicative	hurts, escape, suck, loser, depress, lonely, sad, alone, weak, useless, worth, hate, uncomfortable, pressure, conversation, suicidal, imbalance, myself, blame, torture, safe, worry, intimidate, therapy, medication, break, mine, painful, nobody, shit, life, problems, unsuccessful
2	Non-depressive	football, favorite, sleepy, life, amazing, tips, beautiful, romantic, work, play, perfect, good, weekend, love, smile, hello, friends, follow, meet, team, movie, excited, great, night, brilliant, bored, awesome

4.2   Term Weighting

Term weighting is generally used for capturing the relevance of document terms in information retrieval and text mining applications. We evaluate the relative relevance of a word in relation to a specific document and across the entire corpus using term frequency and inverse document frequency (TF-IDF). Importance of a specific term/word directly depends on how many time a word appear inside the document. It can be understood by taking some example, in case of vectorization, we are giving equal preference to every word whether it is '*the*' or '*torture*', but we can easily say that word '*the*' don't have any importance in depression classification and it is present in most of the tweet so TF-IDF is used to penalize these kinds of words.

The number of times a word appears in the document divided by the total number of words in the document yields TF (Eq. 6). Every tweet is a document for our present work, and the occurrence of terms in a tweet is considered for this calculation. The relevance of an uncommon word in the corpus is captured by inverse document frequency (Eq. 7).

$$TF(t) = a/b \tag{6}$$

where, $TF(t)$ is the term frequency of term t, a is the frequency of particular word in the current document, and b is the sum of frequencies of all words in the corpus.

$$IDF(t) = log_e(TN/ND). \tag{7}$$

TN denotes the total documents and ND is the number of documents that contain a specific term t.

$$TFIDF(t) = TF(t) * IDF(t) \tag{8}$$

4.3 Measuring Depression

For this task, we adopted a standard metric called SMDI (Social Media Depression Index) [33] SMDI is used to measure the level of depression in the individual or population using statistical analysis of influencing factors. It is the standardized difference between the depression post and normal post in the given time t, where t may day, week, etc. In case of individuals, we take the tweets of the social media user, but in case of population, the entire tweet corpus is considered for computation of SMDI (Eq. 9).

$$SMDI = \frac{nd(t) - \mu d}{\sigma d} - \frac{ns(t) - \mu s}{\sigma s} \tag{9}$$

where, $nd(t)$ denotes the number of depression tweets, $ns(t)$ denotes total normal or standard tweets, μ_d is mean of numbers of depression tweets, μ_s is mean of numbers of non-depression tweets, σ_d is the standard deviation of numbers of depression tweets and σ_s is the standard deviation of numbers of non-depression tweets.

4.4 Prediction Models

For the task of depression detection and prediction, we employed an ensemble deep neural model consisting of Convolution neural network with LSTM. The model's architecture is seen in Fig. 1. The model uses a word embedding layer employing Word2Vec [34], which is a. neural embedding model trained on GoogleNews-vectors-negative300 dataset [35]. The first layer is a convolution layer using a liner filter, while the second layer is a max pooling layer, which is used to extract the important semantic knowledge, as all terms do not contribute equally important to the prediction task. The dropout layer is used to avoid overfitting challenges, in our work, we employed 20% dropout.

The LSTM is a recurrent neural network model that attempts to model past knowledge by 'remembering' it and 'forgetting' the information which is not relevant. There are different activation layers called gates for remembering the previous information and forgetting the irrelevant data. A Internal cell state vector is maintained in every LSTM recurrent unit. There are 4 gates in LSTM for different purpose—*Forget gate* (tells what information to be forgotten), *Input Gate* (writes the information to internal cell state vector), *Input Modulation Gate* (a subpart of input gate utilized to alter the information that the input gate writes in the internal cell state by introducing non-linearity and making information Zero mean, allowing for quicker convergence and reduces learning time.) and *Output Gate* (output is generated based on the information stored in the Internal cell state vector). After the LSTM layer, another Dropout layer is used, with a 20% dropout value. The final dense layer is set with the activation function as sigmoid, and the output is a two-class prediction, either 0 and 1. Here,
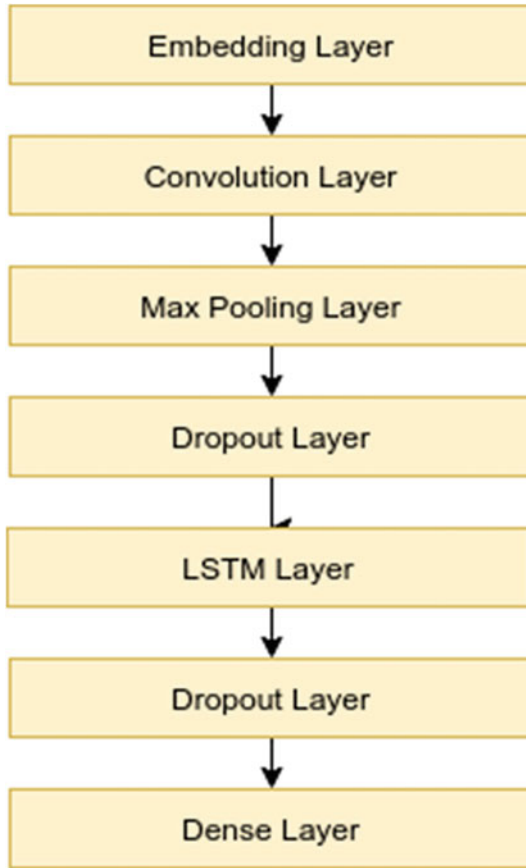
Fig. 1 Architecture of the Depression prediction CNN model

1 denotes the positive class, which indicates that a particular user shows depressive tendencies while 0 denotes that there is no evidence of depression in a given user.

5   Experimental Results and Discussion

We used Sentiment140 dataset which consists of around 1,600,000 tweets. The dataset contains 6 fields i.e. target, id, date, flag, user, and text. The target field indicates if a given tweet is negative, neutral or positive. Here, positive means the person is in depression, negative means the person is happy, and neutral indicates that the person is in normal condition. We trained and tested the proposed model on Google Colab using GPU processor.
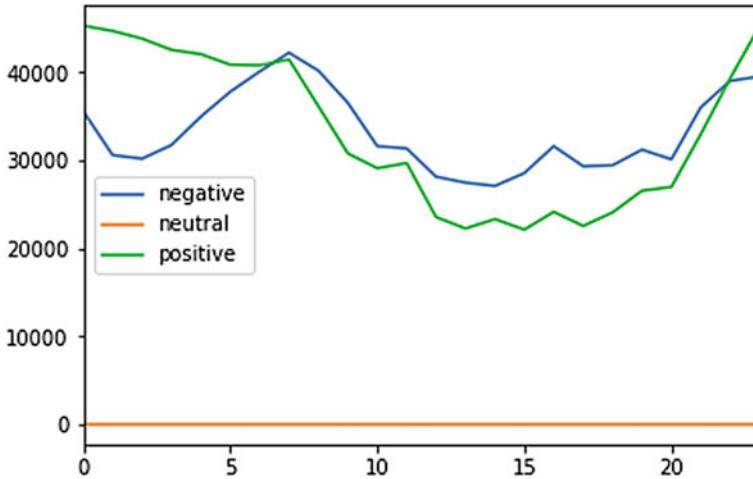
Fig. 2 Analysis of depression with time

5.1 Effect of Depression over Time

Time of social communication is an important factor in detection of depressive tendencies. To evaluate this aspect, we used the time stamp value of the tweet during the analysis, to determine whether the user is depressed or not. For this work, we have defined a day and night window as follows: Day window (6:00–21:00) and Night window (21:00–6:00). The results of this analysis are shown in Fig. 2, where it can be seen that the black line (Night window) is almost always above the black line (Day window). Based on this, we could conclude that the prevalence of depressive behavior is more at night time when compared to day because of feelings of loneliness and aloofness when alone at night time.

5.2 N-Gram Analysis Using SVM

Next, we conducted experiments to understand the contribution of terms used in the tweets by users, toward expression of depressive tendencies. For this, we considered various n-gram feature approaches for representing the tweet text. A contiguous sequence of *n* items retrieved from a text, dataset and speech is known as an n-gram. For our experiments, we considered different types of n-gram representations—unigrams, bigrams, trigrams, 4-grams, and 5-grams. Each of these characteristics is taken from user posts in order to account for normal language usage and to model the context. For the prediction, the SVM model trained on TF-IDF features is considered. Table 2 presents the observed performance of the model using various n-gram. We can observe that, the bi-gram representation outperformed all other models by a

Table 2 Observed performance of SVM classifier with TF-IDF and n-gram term representations

Features	Accuracy
Uni-gram	0.7075
Bi-gram	0.7200
Tri-gram	0.7085
Four-gram	0.7095
Five-gram	0.7090

Table 3 Effect of TF-IDF on performance of various classifiers

Approach	Classifier	Accuracy
No tf-idf	SVM	0.6905
	Random forests	0.6835
	XGB	0.6985
With tf-idf	SVM	0.7075
	Random forests	0.6985
	XGB	0.7092

small margin, while uni-gram, tri-gram, four-gram and five-gram also performed adequately.

5.3 Comparative Analysis of Three Classifiers

Comparative analysis of the performance of n-gram representations with other classifiers and when TF-IDF was not used, is presented in Table 3. First, we tested all three classifiers without TF-IDF then all three classifiers with TF-IDF. From the tabulated results, it is evident that TF-IDF increased the prediction accuracy of the SVM, while the XGB classifier achieved the best-in-class accuracy.

SVM [36] finds the Hyperplane in such a way that the margin between classes is maximum while the other algorithms, i.e., XGB, Random Forest don't care about the margin, that's why the accuracy of SVM is better as compared to XGB, Random-Forest.

We also used other metrics for assessing the goodness of the prediction of the classifiers. For this, precision and recall metrics were used. Precision is given by the number of true positives divided by the total number of true positives and false positives, where T_P denotes true positive, F_N denotes the false negative and F_P denotes false positive. The recall is calculated by dividing the total number of true positives by the total number of true positives and false negatives. After calculating the precision and recall, we can combine them to get F1-score. This is often used when there the data is imbalanced.

Table 4 Precision, Recall, and F-score performance for SVM classifier

| Method	Class	Precision	Recall	F1-score
Uni-gram	Non-depressive	0.72	0.67	0.70
	Depressive	0.69	0.74	0.72
Bi-gram	Non-depressive	0.73	0.66	0.69
	Depressive	0.69	0.76	0.72
Tri-gram	Non-depressive	0.73	0.65	0.69
	Depressive	0.69	0.76	0.72

$$P = \frac{T_P}{T_P + F_P} \tag{10}$$

$$R = \frac{T_P}{T_P + F_N} \tag{11}$$

$$F1 - Score = \frac{2 * P * R}{P + R} \tag{12}$$

In terms of accuracy, recall, and F1-score, Table 4 shows the depression prediction performance using SVM for various n-gram encoding methods. When compared to the uni-gram representation, the bi-gram and tri-gram have greater accuracy, recall, and F1-score.

The ROC curve was also plotted to observe the Area Under the Curve (AUC). The larger the AUC, the better the model. ROC curve is shown in Fig. 3. The comparative
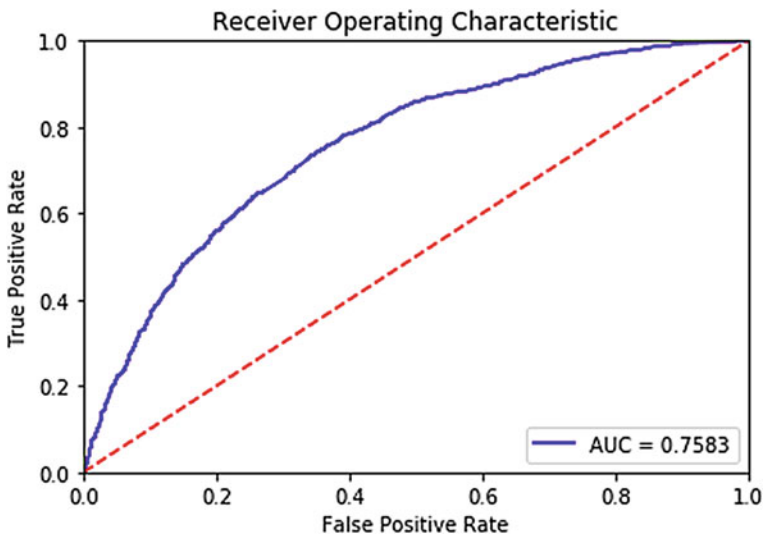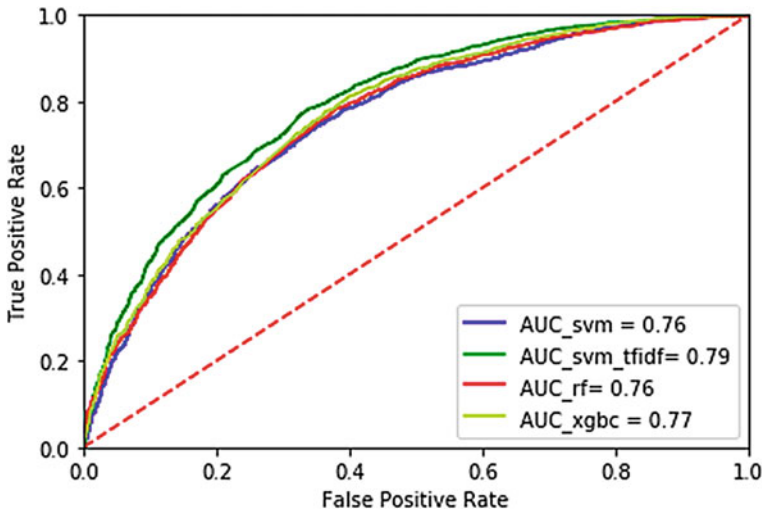

Fig. 3 ROC performance for SVM

**Fig. 4** Comparative AUC obtained for various models

ROC performance for the various classifiers is shown in Fig. 4. The PRC curve is plotted as a function of precision and recall. AUC determines the area under curve, the greater the value of the AUC, the better is the model. It can be seen that the black line has the highest area under curve, thus proving that SVM worked well for all model variants.

Most of the machine learning algorithms like SVM, XGB, Random Forest don't give good accuracy on complex dataset, therefore, we moved to deep learning algorithms like CNN, LSTM, etc., because they are able to learn complex structure inside the data, and hence give better accuracy as compared to machine learning algorithms.

5.4 CNN-LSTM Ensemble Model

During experiments, it was observed that the CNN-LSTM ensemble model outperformed SVM and all other ML models, with an accuracy of 97.1%. Figure 5 depicts the changes in testing and validation accuracy, it can be seen that as the number of epochs increases, both testing and validation accuracy also increases. Table 5 shows the CNN-LSTM model's observed performance in terms of accuracy, recall, and F1-score. When compared to the other machine learning models, the CNN-LSTM model achieved the greatest Precision, Recall, and F1-Score values. This shows that the deep learning model worked better when compared to machine learning models for the depression detection task.

To measure the depression in the population and for assessing the percentage of the population showing depressive tendencies, we used the SMDI (Social media
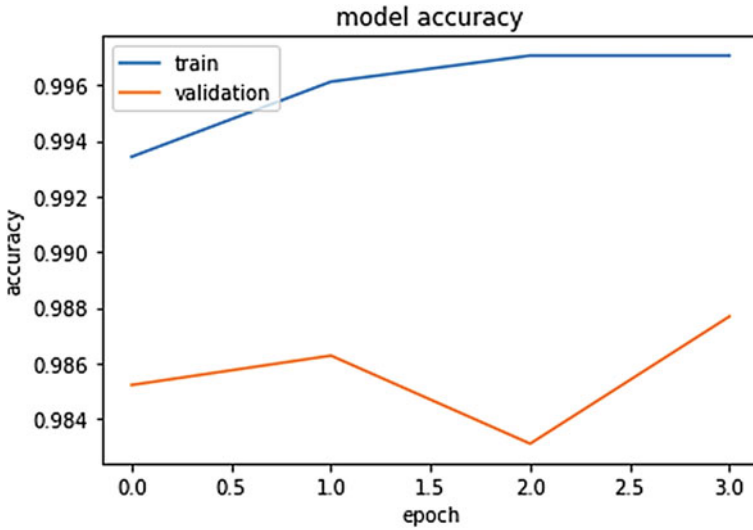
Fig. 5 Training and validation accuracy of the CNN-LSTM ensemble model

Table 5 Accuracy obtained with the CNN model

Class	Precision	Recall	F1-score	Support
0	0.97	1.00	0.99	2379
1	0.97	0.94	0.97	462
avg	0.97	0.97	0.98	2841

depression index) metric. The higher the computed value of this metric, the higher the prevalence of depression in the population. The overall SMDI value for the considered dataset was found to be -136.331764, which is quite low, which in turn shows that only a minority of the considered population exhibits depressive tendencies. Figure 6 shows a plot of the computed SMDI value during daytime and nighttime, it can be clearly seen that the SMDI value is higher for nighttime posts.
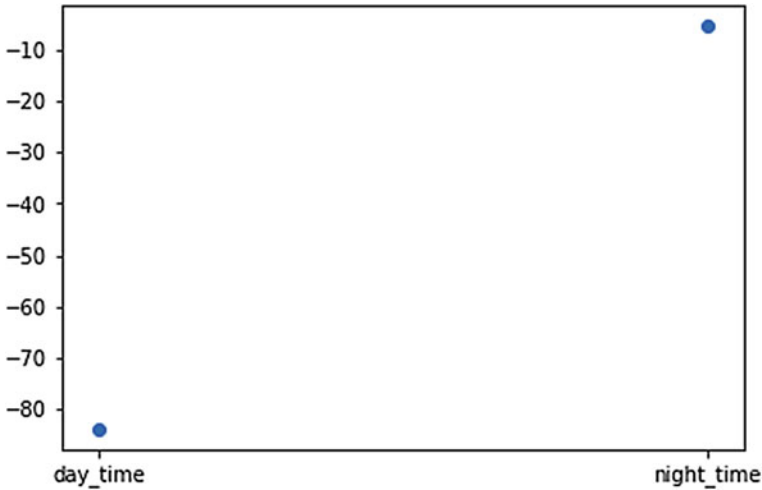
Fig. 6 SMDI value at day and night

6 Conclusion and Future Work

In this work, we presented an approach leveraging ML and DL algorithms for detecting and predicting the prevalence of depressive tendencies among social media users. The Sentiment140 dataset containing more than 1,600,000 tweets was used for the experiments. Term weighting and n-gram modeling approaches were adopted for representing the tweet corpus for feeding into the ML models. SVM, Random forests, XGB, and CNN were applied to the processed and modeled dataset. Experiments revealed that SVM model outperformed all other models when TF-IDF was used, as compared to count vectorization and with bi-gram approaches. The CNN and LSTM ensemble model outperformed all ML models and achieved 97% accuracy. Computation of SMDI metric for the tweet dataset showed that the percentage of depressed persons in the population is low, and the prevalence of depressive behavior is more at night when compared to daytime. In future, we intend to extend our work by using recurrent neural network for better accuracy and also adapt our approach for other social media platforms.

References

1. Depression in population according to WHO. https://www.who.int/news-room/fact-sheets/detail/depression
2. Suicide, prevention and control Strategy by WHO. https://www.who.int/news-room/fact-sheets/detail/suicide

3. Haselton MG, Nettle D, Murray DR (2005) The evolution of cognitive bias. The handbook of evolutionary psychology
4. Resnik P, Armstrong W, Claudino L, Nguyen T, Nguyen V-A, Boyd-Graber J (2015) Beyond lda: exploring supervised topic modeling for depression-related language in twitter. In: Proceedings of the 2nd workshop on computational linguistics and clinical psychology: from linguistic signal to clinical reality
5. Mitchell M, Hollingshead K, Coppersmith G (2015) Quantifying the language of schizophrenia in social media. In: Proceedings of the 2015 annual conference of the north American chapter of the ACL: human language technologies (NAACL HLT)
6. Benton A, Mitchell M, Hovy D (2017) Multitask learning for mental health conditions with limited social media data. EACL
7. Harman GCMDC (2014) Quantifying mental health signals in twitter. ACL 2014
8. Yazdavar AH, Al-Olimat HS, Banerjee T, Thirunarayan K, Sheth AP (2016) Analyzing clinical depressive symptoms in twitter
9. De Choudhury M, Gamon M, Counts S, Horvitz E, Predicting depression via social media. In: ICWSM
10. Nguyen T, Phung D, Dao B, Venkatesh S, Berk M (2014) Affective and content analysis of online depression communities. IEEE Trans Affect Comput 5(3):217–226
11. Coppersmith G, Dredze M, Harman C, Holling- shead K, Mitchell M (2015) Clpsych 2015 shared task:Depression and ptsd on twitter. In: Proceedings of the 2nd workshop on computational linguistics and clinical psychology
12. Coppersmith G, Dredze M, Harman C, Holling- shead K (2015) From adhd to sad: analyzing the language of mental health on twitter through self-reported diagnoses. In: NAACL HLT
13. Neuman Y, Cohen Y, Assaf D, Kedma G (2012) Proactive Screening for depression through metaphorical and automatic text analysis. Artif Intell Med 56(1):19–25
14. Karmen C, Hsiung RC, Wetter T (2015) Screening internet forum participants for depression symptoms by assembling and enhancing multiple nlp methods. In: Computer methods and programs in biomedicine
15. Pravalika A, Oza V, Meghana NP, Kamath SS (2017) Domain-specific sentiment analysis approaches for code-mixed social network data. In: 2017 8th international conference on computing, communication and networking technologies (ICCCNT). IEEE, pp 1-6
16. Ashok M, Rajanna S, Joshi PV, Kamath S (2016) A personalized recommender system using machine learning based sentiment analysis over social data. In: 2016 Ieee students' conference on electrical, electronics and computer science (SCEECS). IEEE, pp 1–6
17. Kaushik R, Chandra SA, Mallya D, Chaitanya JNVK, Kamath SS (2016) Sociopedia: an interactive system for event detection and trend analysis for twitter data. In: Proceedings of 3rd international conference on advanced computing, networking and informatics. Springer, New Delhi, pp 63–70
18. Chandra A, Mallya D, Chaitanya JNVK, Kamath S (2015) Ontology based approach for event detection in twitter datastreams. In: 2015 Ieee region 10 symposium. IEEE, pp 74–77
19. Holleran SE (2010) The early detection of depression from social networking sites. The University of Arizona, Tucson
20. Greenberg LS (2017) Emotion-focused therapy of depression. Per Centered Exp Psychother. 16(1):106–17
21. Haberler G (2017) Prosperity and depression: a theoretical analysis of cyclical movements. Routledge, London
22. Guntuku SC et al (2017) Detecting depression and mental illness on social media: an integrative review. Curr Opin Behav Sci. 18:43–9
23. De Choudhury M, Counts S, Horvitz E (2013) Predicting postpartum changes in emotion and behavior via social media. In: Proceedings of the SIGCHI conference on human factors in computing systems. ACM, New York
24. O'Dea B et al (2015) Detecting suicidality on Twitter. Internet Interv 2(2):183–8
25. Zhang L et al (2014) Using linguistic features to estimate suicide probability of Chinese microblog users. In: International conference on human centered computing. Springer, Berlin

26. Paul MJ, Dredze M (2011) You are what you tweet: analyzing twitter for public health. In: Proceedings of ICWSM '11
27. Sadilek A, Kautz H, Silenzio V (2012) Modeling spread of disease from social interactions. In: Proceedings of ICSWM '11
28. Billings A, Moos Rudolf H (1984) Coping, stress, and social resources among adults with unipolar depression. J Personal Soc Psychol 46(4):877–891
29. Aldarwish MM, Ahmed HF (2017) Predicting depression levels using social media posts. In: 2017 Ieee 13th international symposium on autonomous decentralized systems
30. Hussain J et al (2015) SNS based predictive model for depression. In: Geissbühler A, Demon-geot J, Mokhtari M, Abdulrazak B, Aloulou H (eds) ICOST 2015, vol 9102. LNCS. Springer, Cham, pp 349–354
31. Biradar A, Totad S (2019) Detecting depression in social media posts using machine learning. In: Recent trends in image processing and pattern recognition. RTIP2R 2018. Communications in Computer and Information Science, Solapur
32. Mowery DL, Park A, Bryan C, Conway M (2016) Towards automatically classifying depressive symptoms from Twitter data for population health'. In: Towards automatically classifying depressive symptoms from Twitter data for population health, pp 182–191
33. Social media as a measurement tool of depression in populations. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.294.8952rep=rep1type=pdf
34. Distributed Representations of Words and Phrases and their Compositionality. https://arxiv.org/pdf/1310.4546.pdf
35. GoogleNews-vectors-negative300. https://code.google.com/archive/p/word2vec/
36. Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20(3):273–297