Generating a Mental Health Curve for Monitoring Depression in Real Time by Incorporating Multimodal Feature Analysis Through Social Media Interactions

Moumita Chatterjee, Aliah University, India
Piyush Kumar, Accenture Services Pvt. Ltd., Kolkata, India
Dhrubasish Sarkar, Supreme Institute of Management and Technology, India*

ABSTRACT
The coronavirus pandemic has led to a dramatic increase in depression cases worldwide. Several people are utilizing social media to share their depression or suicidal thoughts. Thus, the major goal of the proposed study is to examine Twitter posts by users and identify features that may indicate depressed symptoms among online users. A numerical metric for each user is proposed based on the sentiment value of their tweets, and it is demonstrated that this feature can detect depression with good accuracy by using several machine learning classifiers. The paper proposes a novel method for measuring the mental health index of an individual by combining the sentiment score with multimodal features extracted from his online activities. A real-time curve is generated using this index that can monitor a person’s mental health in real time and offer real-time information about his state. The proposed model shows an accuracy of 89% using SVM, and proper feature selection is very essential for obtaining good performance.

KEYWORDS
Depression, Mental Health Curve, Natural Language Processing, Real Time, Social Media

1. INTRODUCTION
Mental health is a severe public health concern across the world due to its high incidence, as well as the suffering, dysfunction, morbidity and the financial burden that it leaves behind. The World Health Organization (WHO) estimated that active treatment for mental health condition is required by around 15% of persons in India (Who.int., 2022). Idiopathic Development Intellectual Disability

DOI: 10.4018/IJIIT.324600

*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.
(4.5 percent) is the most frequent type of mental condition in India, followed by Depressive Disorders (3.3 percent) and Anxiety Disorders (3.3 percent) (Statista, 2022). Lancet research states a major correlation exists between suicide rate and depressive diseases in India at the state level (Sagar et al., 2020). The main symptoms of depression include chronic melancholy, loss of interest in previously-enjoyed activities, and an inability to function normally for at least two weeks (Zimmerman et al., 1987). As a result, successful therapy for depression depends on a timely diagnosis (Ríssola et al., 2020). However, because of the social stigma associated with depressed symptoms, many people do not seek professional aid or therapeutic guidance (Zou et al., 2020). So, in order to address their difficulties, individuals are turning to informal sources like social media.

Since the advent of social media, many who struggle with mental health concerns has found comfort in sharing their thoughts and experiences in online forums, tweets, and blogs (Park et al., 2012; Bathina et al., 2021). According to psychological research, there is a substantial link between an individual’s emotional well-being and their language use. Today, Twitter is the most widely used social media platform for sentiment analysis research. Researchers believe that analysis of Twitter messages may be used to detect depression and other mental health issues. These online activities inspire them to create new methods for the early diagnosis of depression and for developing potential health care solutions. This is accomplished by using machine learning algorithms in conjunction with Natural Language Processing (NLP) approaches to detect depression in user posts.

Although significant progress has already been achieved in this subject, there are still certain obstacles to overcome. In our previous studies (Chatterjee et al., 2022a; Kumar et al., 2022; Chatterjee et al., 2022b; Samanta et al., 2022; Sarkar et al., 2022; Sarkar et al., 2023), we aimed to attain a high performance in the early depression diagnosis by carefully selecting features from Twitter tweets. In this study, we intend to extend our prior research on depression detection by aiming to continually track a person’s mental health and provide real-time information about his condition. While a large body of work has provided models for real time depression detection and early depression detection (Angskun et al., 2022; Babu et al., 2022; Chenhao et al., 2020), to the best of our knowledge, there are no works that provide models for continuous tracking of a person’s mood over a period of time to understand the sentiment dynamics of the individual. A depressed person needs constant real-time monitoring since someone who is at risk of depression could not be conscious of his own activities. If the individual’s curve exhibits a downward tendency (negative sentiment) for a prolonged period of time, an alarm can be raised in advance in case of emergency. This effort primarily intends to accurately model each individual’s sentiment in the form of a mental health curve that can be utilized to reproduce the sentiment dynamics of tweets.

For plotting the mental health curve from the user’s social media interaction, a number of NLP and machine learning techniques were applied. This novel study aims to investigate continuous real-time depression detection from user posts on social media. In order to understand depression from user posts, this work comprises a thorough examination of language preferences, emoticon usage, posting time, and subject descriptors. Four learning algorithms were used to predict depressive thoughts using a variety of factors that were extracted from the data. Additionally, statistical studies of the posts were conducted in connection to the extracted features, and several intriguing facts, such as the average length of tweets, the frequency of emoticons, and language usage, were found. Depressive and nondepressive writings usually cover a diverse range of topics, allowing us to better comprehend the two types. Latent Dirichlet Allocation is used to extract a set of latent topics from writings that are both normal and depressive. In order to achieve this goal, an effort has been made to determine a subset of personality traits utilising significant features that may serve as a depression risk indicator. Furthermore, we developed a measure for Sentiment Analysis of user posts. Finally, to identify depressive ideas in user postings, these feature sets are integrated with classification techniques. The effectiveness of each feature individually and in various combinations is examined.

Our main contributions are as follows:
1. Linguistic, Topic, Emoticon, and Sentiment features are chosen for the study challenge and the depression detection performance of each feature type and their numerous combinations are shown.

2. A polarity-preserving numerical rating is proposed based on the emotional score of each user’s tweets.

3. A normalised numerical score is assigned to each of the features extracted from user posts based on the depression severity level that can be deduced from the data.

4. A metric is proposed for integrating multimodal information and visualising the results in the form of a graph that represents an individual’s mental health over time.

5. Machine learning approaches are used to demonstrate the advantages of combining various features for achieving a greater accuracy in depression detection.

The rest of the paper is organized as follows. Section 2 consists of the literature review on depression detection. The proposed methodology is described in Section 3. Section 4 discusses the experimental results. Discussions and Conclusions are given in Section 5 and 6.

2. LITERATURE REVIEW

This section aims to offer a comprehensive overview of the many studies linked to the identification of mental illness on social networks.

Several researchers extracted single set feature groups such as N-grams (Wongkoblap et al., 2017), Bag-of-Words (Benton et al., 2017; Nadeem, 2016), LIWC (Paul et al., 2018) or LDA (Coppersmith et al., 2015; Maupomé et al., 2018) for diagnosing depression in user postings. Other studies (Resnik et al., 2015; Preoţiuc-Pietro et al., 2015; Nguyen et al., 2014; Schwartz et al., 2014; Tsugawa et al., 2015) compared the effectiveness of each of these distinct characteristics using different machine learning techniques. Some recent research projects have concentrated on boosting detection accuracy by merging some of the features. Wolohan et al. (2018) used N-Gram and LIWC to increase the detection accuracy across a single set of features.

Tyshchenko et al. (2018) integrated N-Gram+LIWC to increase the detection accuracy compared to a single group of features. Tadesse et al. (2019) employed a similar sophisticated text pre-processing method and combined TF-IDF, LDA and Bag of Words with Convolutional Neural Networks (CNN) to enhance performance. Benamara et al. (2018) discovered that combining features can lead to increased performance. They evaluated the efficiency of a single feature, such as a bi-gram, with a Support Vector Machine (SVM) classifier to achieve an accuracy of 80 percent. They also showed the effectiveness of combining features (LIWC+ LDA+ bi-gram) with Multilayer Perceptron (MLP) to get 91 percent accuracy. A substantial amount of work has emerged in recent years that discusses how to identify mental illness early by using data from social media (Song et al., 2018; Cacheda et al., 2019; Sarkar et al., 2019).

De Choudhury et al. (2013) identified a significant but under-reported mental health issue among a large number of females. They focussed on using Twitter tweets to develop a prediction model of the impending impact of delivery on the disposition and conduct of new moms. For identifying postpartum changes, 376 moms’ Twitter postings were evaluated along features such as linguistic style, social networking activities, emotions, and social network utilization (Dutta et al., 2021). De Choudhury et al. (2013) used both human coders and a machine learning classifier to verify whether the text of the suicide-related tweet will be able to depict the degree of anxiety.

According to O’dea et al. (2015), individuals are increasingly using social media to talk about their emotions and feelings. They created a well-labeled data set on depression using Twitter postings and extracted six feature groups composed of depression-related features from online social behaviours clinical data. The feature groups were used to construct a multimodal depression dictionary learning model for identifying depressed Twitter users. Shen et al. (2017) conducted extensive research on the link between mental health and social media. They linked anxiety-depression to erratic thinking, irritability, and sleeplessness. They suggested a real-time
twitter prediction algorithm for anxious depression utilising posting patterns and language clues. An ensemble voting classifier is used to execute majority voting on the feature vector, which is made up of emotion, timing, frequency, contrast, and the word. Kumar et al. (2019) used Twitter tweets to categorise positive and negative emotions of users and compared the findings to deep learning approaches. CNN-based deep learning surpasses machine learning models by a large margin, according to the authors. Lora et al. (2020) also used Twitter tweets to create depression training and test datasets, and the classifier is Naive Bayes.

According to the authors’ (Rao et al., 2020) finding, with the increased usage of social media, individuals are talking about their depression, particularly suicide thoughts, on social media. They examined several Twitter profiles and employed several accounts and tweet-related variables to predict suicidal profiles. The approach’s efficacy was determined using a dataset of persons who had previously committed suicide. In the same vein, Mbarek et al. (2019) highlights a recent research in which the scientists investigated classification tasks to diagnose depression by analysing text messages from Reddit users. The bigrams, TF - IDF, and embedding were regarded as baseline features.

Leiva et al. (2017) combined a standard surface characteristics, bag of words and linguistic features to build a prediction model for detecting early signs of depression. Stankevich et al. (2018) suggested a Feature Attention network comprised of a series of feature networks spanning from depression symptoms, attitudes, ruminative thinking, and writing style and utilised deep learning to detect depressed individuals.

Geetha et al. (2020) investigated several strategies for early detection of depression using single and dual machine learning techniques. Tadesse et al. (2020) utilised two separate approaches for early identification of symptoms: the first examines the user’s time and writing patterns, while the second adds hints from shared text and tweets. Ramírez-Cifuentes et al. (2020) addressed the early identification of suicide using a deep learning LSTM-CNN and machine learning based categorization technique using Reddit social media posts.

Despite the fact that some of the work listed above has examined sentiment, language, and topic analysis for depression identification, there are major gaps in the existing literature. Only a few research have concentrated on using Emotional, Sentiment, Linguistic and topic qualities on their own, but no well-known studies have combined all of these features and applied them to the same dataset to notice the divergence in the findings. We overcome these limitations in our current work by attempting to detect depression from Twitter posts. In addition, to the best of our knowledge, no other works have attempted to build a mental health curve by merging multimodal feature set for real time tracking of a person’s emotions. The goal of this research is to integrate several feature sets to create a mental health index for a person. This index will give a temporal perspective of the individual’s mental condition across time and may be used to alert his parents in an emergency.

3. PROPOSED MODEL

3.1. Real-Time Depression Detection

The goal of this study is to develop a real-time depression detection model by collecting streaming data from Twitter’s API (application programming interface), extracting, converting, and loading data into the data storage. The depression detection model then uses this data as an unobserved input. The model categorises the data in order to determine a person’s level of depression.

The model operates in two modes: suspect mode and parental or guardian mode. In suspect mode, a user can utilise his own Twitter Id as input to the algorithm to determine his own level of depression. In parental mode, parents may follow their child’s status by entering the child’s Twitter ID into the system. Three parameters are needed for the model to work: the Twitter username, the starting date, and the ending date, which indicates the time window for scraping tweets.
Furthermore, the algorithm outputs a curve to represent the temporal change of a person’s emotions based on the sentiments expressed in his tweets. Figure 1 shows the Real time depression detection model. The detailed methodological framework of the model is given in Figure 2.

### 3.2. Methodology

A two-level Depression Profile Detection framework is outlined in this research for detecting depressed profiles on Twitter. Initially, a range of profiles is analyzed, and a variety of criteria are utilized to classify people as depressed or non-depressed. These features can be derived either explicitly from user profiles or implicitly via the use of various machine learning techniques and methodologies. Here, the authors primarily focused on psychological, emotional, and personality trait analysis of each user, which reveal details on the mental health of the suspected user group.

The individuals in the suspect dataset were then exposed to a second classification process in which their tweets were reviewed and different linguistic, temporal, and topic factors were employed to identify persons from the suspect group in order to detect depression signs.

Each user is then represented as a vector that contains all of the features that were used, and the effectiveness of each feature is compared with the effectiveness of the suggested framework. The proposed methodology’s structure, shown in Fig. 2, includes data preprocessing, feature extraction, data storage, analysis, followed by model construction and evaluation and output. The proposed method is outlined in the following sections.

#### 3.2.1. Data Set Exploration and Preparation

The proposed methodology trains the algorithm for depression detection using data from Twitter and Reddit. The first step was to create a compilation of quotes about depression from subreddits where people ask for help from the internet community. These posts are typically posted by persons with depressed tendencies, therefore they can be classified as depressive-indicative remarks. Additionally, normal posts from other subreddits about friends, family, or entertainment are also collected. The depressive-indicative postings were carefully examined to establish a group of words that were then utilized as Twitter search terms. Some of the words used in both sorts of postings are shown in Table 1.

![Figure 1. Real-time depression detection model with depression curve generation](image-url)
A total of 188,704 tweets containing these search phrases was retrieved from 2000 users using Twitter APIs. Among these tweets (37740/188704), 400 persons’ tweets were held aside for testing purposes. The remaining tweets were manually labeled in order to generate two datasets, one for depressed users and one for non-depressed users. Initially, tweets with the highest amount of depressive-indicative terms relating to a certain person were chosen. These tweets were then manually tagged in two data sets to classify users. Users with depression-like behavior are included in the depressive dataset. The non-depressed dataset contains persons whose tweets did not appear to represent depressive thinking, users who did not mention personal problems, and users who reported news or remarks regarding depression. The training dataset statistics is given in Table 2.

For each Twitter user, a simple set of statistics about their profile and tweets were collected during a three-month period beginning with an anchor tweet. The authors used the data and calculated the average amount of hashtags, links, responses, and @mentions per tweet. Instead of integrating all tweets for a certain user, each tweet is treated as a post for a session in this work. By analyzing as few tweets as possible, the proposed approach analyzes each tweet independently in order to determine if a person is depressed or not.

Due to the existence of various noises in the raw data, data obtained from online social media cannot be utilized directly for feature extraction. This makes semantic analysis and word matching more difficult. Data from online social networking sites may contain grammatical and spelling mistakes, emoticons, and other unwanted features, exacerbating the situation. As a result, the data...
must be preprocessed to guarantee that the computer model performs reliable predictive analysis. The data are subjected to the following preprocessing techniques:

- URL links in user posts are eliminated as part of preprocessing since they lack significance or polarity.
- Stop words such as ‘a,’ ‘an,’ ‘the,’ and so on are removed because they are not discriminatory or useful in our model.
- Non-ASCII characters are eliminated to enhance text quality.
- Tokenizing is used to convert sentences into a collection of single words.
- Stemming is the process of converting each word into its root word.
- POS (Part of speech) tagging is used to eliminate ambiguity in word interpretation.

Data preparation and noise reduction removes noisy data material, resulting in a high-quality and dependable dataset that may be utilised in this research. Furthermore, the data preparation stage reduces the computational complexity of the model because the study only has to deal with useful data that will be utilised in the model.

3.2.2. Feature Extraction

Feature extraction is an important step in acquiring thorough information about users in order to accurately diagnose depression. A large variety of factors help to increase the accuracy of depression identification.

3.2.2.1. 1st Level Feature Extraction

The features that are considered in the first level classification are:

1. Social media features

   This feature category contains information collected directly from a user’s profile, such as the user’s name, language, number of friends, profile description, country, profile creation date, time zone, profile photo, number of followers, and followers. These attributes can be used to produce other aspects such as profile polarity and subjectivity. The user profile description’s polarity value represents the individual’s persona and may be utilised as a key component in the process of classification. The social media elements included in the analysis are listed in Table 3.

2. Temporal features

   By examining temporal features, we may gain a better understanding of how individuals share content on Twitter at different times of day, such as in the morning or evening. According to a dataset

<table>
<thead>
<tr>
<th>Description</th>
<th>Depressed Dataset</th>
<th>Non-Depressed Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Twitter users</td>
<td>672</td>
<td>928</td>
</tr>
<tr>
<td>Number of tweets</td>
<td>74125</td>
<td>76839</td>
</tr>
<tr>
<td>Mean number of tweets per user</td>
<td>166.57</td>
<td>158.25</td>
</tr>
<tr>
<td>Mean tweet length</td>
<td>142.89</td>
<td>120.77</td>
</tr>
<tr>
<td>Mean number of emoticons per user</td>
<td>30.84</td>
<td>42.80</td>
</tr>
</tbody>
</table>
analysis, the AM value appears more frequently in the depressed data set than in the non-depressive data set. Because of loneliness, a break from work, a loss of energy, and alterations in communication between light and darkness and the neurological system, depressive and suicidal thoughts are more prevalent at night and in the early morning hours.

According to a depression dataset analysis, around 76 percent of users who had depressive thoughts were active between 12am to 6am and 42 percent were active between 6 am to 12 pm. In the non-depressive sample, only 58% of users were active between 6 p.m. and 6 a.m. According to another study, just 6% of users with depressed thoughts were active between 12pm and 6pm, compared to 21% of non-depressive users. The temporal analysis of depressed and non-depressive users is shown in Figure 3.

In our analysis we have divided the dataset into 4 time zones-12am-6am,6am-12pm,12pm-6pm,6pm-12am. After analysis of both the datasets, it has been observed that user mostly have depressive thoughts during the AM time zone. So this feature is used to compute the number of tweets that the user posted online during each of the time zone.

The metric for the number of posts per time period is defined as follows:

### Table 3. Description of social media features

<table>
<thead>
<tr>
<th>Features</th>
<th>Description</th>
<th>Source of Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>#followers</td>
<td>Followers count</td>
<td>Metadata of tweet</td>
</tr>
<tr>
<td>#following</td>
<td>Number of accounts that the user follows</td>
<td>Metadata of tweet</td>
</tr>
<tr>
<td>#friends</td>
<td>Friends count</td>
<td>Metadata of tweet</td>
</tr>
<tr>
<td>Polarity of profile description</td>
<td>Calculates the polarity of the user profile description and outputs a floating point number between [-1,+1] to indicate whether it is negative or positive.</td>
<td>Using the Python TextBlob function on a profile description derived from Tweet metadata.</td>
</tr>
<tr>
<td>Subjectivity of profile description</td>
<td>The subjective value of the user profile description is calculated. Indicates if the context is subjective or objective by providing a floating point number between [0.0 and 1.0].</td>
<td>Using the Python TextBlob function on a profile description derived from Tweet metadata.</td>
</tr>
</tbody>
</table>

Figure 3. Temporal analysis of depressive and non-depressive datasets
Count of posts/Time Unit ($C_{TZ}^i$): A numerical score ($TZ_{score}^i$) is assigned to each time zone with AM time zone having lower scores and PM time zones having higher scores, where $i$ denotes the day on which the score is calculated. The scores are then normalized to a common scale of [0,1] to ensure uniformity with all the other feature values. The normalized $TZ_{score}^i$ is computed using the following equation:

$$\forall_i \text{Norm}(TZ_{score}^i) = \frac{TZ_{score}^i - \text{Min}(TZ_{score}^i)}{\text{Max}(TZ_{score}^i) - \text{Min}(TZ_{score}^i)}$$  \hspace{1cm} (1)$$

After normalization, the score is multiplied with the count of the posts ($C$) done during that zone.

$$C_{TZ}^i = \sum_i (TZ_{score}^i \times C)$$  \hspace{1cm} (2)$$

3. Emojis/emoticons sentiment

Emojis are used by users to communicate nonverbally and using simple symbols. Emojis can be utilised to stimulate the reader’s curiosity. Emojis may help you understand the sentiment behind any text or tweet, and it’s critical to discern between positive and negative sentiment elements. Tweets from users typically include a plethora of emoticons representing a wide range of emotions such as love, fury, surprise, happiness, sadness, and fear. A user’s emoji use is calculated in this study utilising a range of metrics.

Several facts about the use of emoticons were revealed when analyzing depressed ideation texts. Depressive persons use less emoticons on average, and the number of emoticons used in depressive messages was shown to be substantially lower than in non-depressive postings. The most often used emojis among depressive-tendency users are folded hands, expressionless face, begging face, loudly sobbing face, disappointed face, screaming in terror face, perplexed face, and face without a mouth. Normal users’ favorite emojis are: face with tears of pleasure, loudly sobbing face, rolling on the floor laughing, red heart, clapping hands, winking face, smiling face with heart eyes. Based on this data, it can be stated that depressive-tendency users use negative emoticons more frequently than non-depressive users. We define the metric for emoticon usage as follows:

**Emoticon Usage Index ($E_{index}^i$):** This index assigns high normalized values to positive emojis and low values to negative emojis. The most often used emojis among depressive-tendency users are assigned low scores. Normal emojis are assigned high scores. The scores are then normalized to a common scale of [0,1] to ensure uniformity with all the other feature values.

4. User biography sentiment analysis

Sentiment Analysis is an efficient method for identifying emotional content in user tweets. This strategy works best when the text has subjective information, such as depression. Sentiment analysis categorizes emotions as either positive, negative, or neutral. This function may determine if a person is feeling good, negative, or neutral based on their profile state. SentiWordNet is a Python tool that assigns positivity, negativity, and objectivity scores to each word in a tweet for sentiment analysis.

5. User tweet sentiment analysis
This method is the most effective for assessing the user’s thinking. The user’s tweets are reviewed 
here to ascertain the sort of emotion shown by the person on Twitter. To begin sentiment analysis,
the SentiWordNet Python programme is used to give polarity and subjectivity values to each word 
in a tweet. We define the polarity index as follows:

**Polarity of the Posts** ($\text{Polarity}_{ji}$): This feature has three attributes positive, negative, and neutral.

Here $i$ and $j$ denotes $i$th day and $j$th user respectively. $\text{Polarity}_{ji}$ is computed as follows:

An equation is used to assign a polarity score to each tweet (3).

$$Tw_{Sc}^k = \prod_{l=1}^{n} \frac{Sc(l)}{\text{nwords}}$$  \hspace{1cm} (3)

$Sc(i)$ denotes the positive or negative scores of each word in the tweet, and $\text{nwords}$ is the number 
of words in the tweet. Taking the product of each word’s score allows you to preserve the sign of 
positivity or negativity. Due to the possibility of enormous numbers being produced by multiplying 
the scores, the scores are normalised using equation (4).

$$\forall_k \text{Norm}(Tw_{Sc}^k) = \frac{Tw_{Sc}^k - \text{Min}(Tw_{Sc}^k)}{\text{Max}(Tw_{Sc}^k) - \text{Min}(Tw_{Sc}^k)}$$  \hspace{1cm} (4)

The score for each user is calculated by averaging the normalised scores across all tweets. In 
equation (5), the equation for computing the scores for each user is presented. The value $n$ represents 
the number of tweets sent by each unique user.

$$\text{Polarity}_{ji} = \sum_{k=1}^{n} \frac{\text{Norm}(Tw_{Sc}^k)}{n}$$  \hspace{1cm} (5)

Sentiment analysis of user tweets from both datasets indicates that most users in the depressed 
dataset have a sentiment score between -0.5 and -1, whereas most users in the normal sample have a 
sentiment score above -0.5. Figure 4 depicts a graph showing the frequency of negative tweets used 
by depressed and non-depressive individuals.

The metric for Subjectivity score and frequency of negative posts are computed as follows:

**Subjectivity of the Posts** ($\text{Sub}_{ji}$): This feature has two attributes: Subjective and Objective. 
Subjective posts contain personal information, emotions, judgment of the users while objective posts 
mainly contain factual information. It has been observed that a depressed person mainly posts subjective 
contents online containing their emotions, thoughts, and feelings. Thus, for this feature low scores are 
assigned to subjective posts and high scores are assigned to objective posts.

**Frequency of Negative Posts** ($\text{Freq}_{Neg}^j$): Depressed users generally posts a large number 
of negative posts compared to normal users so this feature can be used for computing the mental health 
index. Similar to 1, this feature is also subdivided into four ranges depending on how frequent an user 
is posting negative comments: $0\%-30\%, 30\%-60\%, 60\%-90\%, >90\%$. The four ranges are assigned a 
numeric score and normalized as above with $>90\%$ getting the lowest score.

6. Personality trait extraction
Every user in the dataset has a unique personality. It represents the numerous personality traits on a particular attribute. From here, it is possible to identify whether the individual is an introvert or an extrovert. The polarity of a user’s tweets, as well as other statistical factors, is computed for each user to determine that user’s personality trait. We examine five key personality traits: agreeability, optimism, activity and sociality, neuroticism, and spectatorship. The polarity of the whole corpus of the suspicious set is first established. The polarity of each user tweet was then compared to the corpus. A person is categorised as optimistic if they have the majority of positive tweets; otherwise, they are categorised as neurotic. Similarly, the polarity of an agreeable user’s tweet is typically the same as the polarity of the corpus as a whole. When the general polarity of the corpus is neutral, a tweet from an active and social individual has a positive polarity. The model calculates a personality score for each user using the following equations:

\[
\text{Active and Social} = \frac{|Positive Posts| + |Negative Posts|}{|Neutral Posts|} \quad (6)
\]

\[
\text{Spectator} = \frac{|Neutral Posts|}{|Positive Posts| + |Negative Posts|} \quad (7)
\]

\[
\text{Pessimist} = \frac{|Negative Posts|}{|Positive Posts|} \quad (8)
\]

\[
\text{Optimist} = \frac{|Positive Posts|}{|Negative Posts|} \quad (9)
\]

\[
\text{Aggregable} = \max \left( \frac{|Negative| + |Neutral| + |Positive|}{|Total Posts|} \right) \quad (10)
\]
Personality Index (PI): Each user is assigned a score for each of the personality trait. After calculation, if any score exceeds 1, then that user is marked as positive for that trait. Analysis of the datasets reveals that user in the depression dataset generally has higher values for Spectator or Pessimist trait and lower values for Optimist trait. Similar to the above scoring method, each personality trait is assigned a normalized score with Pessimist being the lowest score and Active Social is assigned a high normalized score.

3.2.3. 1st Level Classification
A binary classifier is developed at the first level to distinguish between two sorts of people: those who are depressed and those who are not. We defined a threshold value for each characteristic and labelled each user as suspicious if they exceeded it. The various thresholds for the first level features are shown in Table 4.

3.2.4. 2nd Level Feature Extraction
The tweets from the suspicious set and the non-depressed set are analyzed in the second level feature extraction to corroborate the classification achieved in the first level. At the second level, the following features are investigated:

3.2.4.1. Topic Features
Topic distribution, also known as topic modelling, is a statistical modelling approach used to identify abstract themes in a collection of text texts. A topic model is developed using LDA to extract the hidden themes from the depressed dataset. The LDA specifies the amount of topics that must be produced. The amount of subjects given determines the accuracy of categorisation. The authors tested different numbers of topics for the research and determined that 10 was a good quantity. By picking 10 themes and integrating them with additional characteristics using SVM, an accuracy score of 89 percent is obtained.

3.2.4.2. N-Gram Modeling and Tf-Idf
The Scikit-learn Python module’s Tf-Idf vectorizer is used to extract unigrams and bigrams. The stop words were removed from the dataset, and the term-document matrix was reduced to the most common unigrams and bigrams. Our complete training dataset is categorised in order to distinguish between depressed and standard message lexicons. The frequencies of all bigrams and unigrams are computed for each post category, and the top 100 unigrams and bigrams for each category are chosen. Negative emotions, feelings, self-obsession, suicidal thoughts, anger, wrath, negative word, despair, interpersonal processes, meaninglessness, and the present tense are lexical items that are predictive of depression, according to the research. Depressive postings may also contain lexicons regarding physical problems such as fatigue, sleeplessness, poor energy, or hyperactivity. Regular

Table 4. 1st level classification features and their threshold values

<table>
<thead>
<tr>
<th>Features</th>
<th>Threshold for Suspected User Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biography polarity</td>
<td>&lt;0</td>
</tr>
<tr>
<td>Temporal features</td>
<td>Posting time at night(00.00 to 05.00)hr (AM)</td>
</tr>
<tr>
<td>Emoticon sentiment</td>
<td>Percentage of negative emoticons used &gt; 70%</td>
</tr>
<tr>
<td>Tweet Sentiment</td>
<td>Negative value</td>
</tr>
<tr>
<td>Biography sentiment</td>
<td>Negative value</td>
</tr>
<tr>
<td>Personality trait</td>
<td>Neuroticism &gt; 1.</td>
</tr>
</tbody>
</table>
post lexicons, on the other hand, include terminology describing previous events, social interactions, and family-oriented words.

3.2.4.3. Tweet Statistics

This feature category includes statistical metrics collected from user tweets. Some communications use simple, short language, while others use complicated sentences and lengthy paragraphs. The number of tweets and their durations, the number of depression-related tweets per user, and the ratio of depression-related postings to the total number of posts per user are among the metrics considered. The average length of tweets for depressed and non-depressive individuals are depicted in Figure 5. According to the findings, users with depressed intentions had significantly higher average text lengths than regular users. This is because users suffering from severe depression are more likely to have a mental illness or social issues, which are reflected in their communications.

Mean Length of Posts ($Length_{mean}$): This feature is subdivided into three categories: 0-50 words, 50-100 words and 100+ words with the first category being assigned a high normalized value and the last category being assigned a lower normalized value.

3.2.4.4. Other Features

It is obvious that each social media user has a particular writing style. Some individuals, however, may utilise a similar writing style while addressing related topics. For example, while posting about emotional topics, people may employ certain language qualities such as elongation, adjectives, exclamations, and so on. As a result, mining these characteristics may significantly improve the probability of spotting sad consumers. Other aspects include the usage of special characters (i.e. @, $, &,, %, (,, ), +,-, *, /,, =, > etc), the percentage of used hashtags, the word length and sentence length utilised, the percentage of used elongations (e.g., ohhhhhh, noooooo, etc), and the number of frequent terms repeated more than five times. To compute the values of these attributes, a python method is created that uses the tweets gathered to determine the values.

3.2.5. 2nd Level Classification

The performance of each type of second level feature and their combinations in accurately identifying depressed users is determined at this level. In identifying depressed characteristics, the performance of the first and second level features are evaluated. A vector including all first and second level features...
characteristics is used to represent each user. Furthermore, the outcome of the first-level classification is handled as a second-level feature. The remaining 20% of the dataset is utilised for testing while the remaining 80% is used for training. The tweets in the collection are divided into social media sessions, with each session containing all of the tweets from that session. Each tweet is labelled as either depressed or non-depressive. Logistic Regression, Random Forest, Support Vector Machine, and XGBoost are among the classifiers used in the model’s design.

4. RESULT AND ANALYSIS

4.1. Mental Health Curve Generation

This section aims to generate a real time curve derived from user posts on social media by combining multimodal features. The curve will be able to provide a temporal analysis of the dynamic changes of an user’s emotions. Section 4.1.1 depicts a single mode feature – sentiment analysis that is used to derive a curve showing the subjectivity and the polarity of an user posts. The mental health curve is generated by combining multiple features with this single mode feature.

4.1.1. Sentiment Curve

We use Sentiment Analysis to develop two curves depicting the subjectivity and the polarity of the user tweets.

The mean subjectivity score of each user is determined over a seven-month period and plotted against his tweet scores to examine the variability in his subjectivity score. Figure 6 and 7 shows a graph for a user’s subjectivity score from the depressed dataset and non depressed dataset respectively.

According to the figure, the individual in question uses social media mostly to convey his own personal beliefs, ideas, feelings, or thoughts. Similar to this, the majority of the users in the depressed dataset have high mean subjectivity scores determined over a seven-month period. Observing the curve on a daily basis allows us to gain insight into the person’s inner thoughts and how he expresses himself while interacting with others.

Figure 6. Subjectivity graph of a depressed user plotted against mean subjectivity over 7 months
The polarity value of a sentence lies between \([-1, 1]\) where -1 is a negative statement and 1 means a positive statement. The mean polarity score of each user is determined over a seven-month period and plotted against his most recent score calculated using equation (3) to examine the variability in his polarity score. Figure 8 and 9 shows a graph for a user’s polarity score from the depressed and non-depressed dataset respectively:

The subjectivity and the polarity curve gives us an idea about the behavior of each user in terms of their sentiment and the feelings that they share on social media.
4.1.2. Mental Health Curve

We now define the Mental Health Index by combining multimodal features with the sentiment and polarity scores for plotting the Mental health curve.

**Mental Health Index (MHI\textsubscript{SM}):** The Mental Health Index (MHI\textsubscript{SM}) value of a user gives a measure of the mental health status computed through the use of various features extracted from the user posts on social media. The features used for computing the index are – Count of posts/Time unit, Frequency of negative posts, Personality Index, Emoticon usage index, Mean length of posts, Polarity of the posts and the subjectivity of the posts.

The multimodal function defining $MHI\textsubscript{SM}\textsuperscript{i}$ is as follows:

$$MHI\textsubscript{SM}\textsuperscript{i} = F(C\textsubscript{TZ}\textsuperscript{i}, Freq\textsubscript{Neg}\textsuperscript{i}, PI, E\textsubscript{index}\textsuperscript{i}, Length\textsuperscript{i}_{mean}, Polarity\textsuperscript{i}, Sub\textsuperscript{i})$$  \hspace{1cm} (11)

Where $F = \prod x\textsuperscript{j}$ where $x\textsuperscript{j}$ denotes feature for user $j$ on day $i$.

Using equation (11) the $MHI\textsubscript{SM}\textsuperscript{i}$ value of an individual of the depressed dataset is plotted for a period of 1 week to see the changes in his $MHI\textsubscript{SM}\textsuperscript{i}$ index. The red line shows the borderline value and all users having $MHI\textsubscript{SM}\textsuperscript{i}$ below the borderline can be considered as depressed and all users having $MHI\textsubscript{SM}\textsuperscript{i}$ above the borderline can be considered as normal. The borderline function is computed by taking the midrange values of all the features. Figures 10 and 11 show the $MHI\textsubscript{SM}\textsuperscript{i}$ values for a depressed and normal individual respectively for a period of one week. From the figure we can see that the $MHI\textsubscript{SM}\textsuperscript{i}$ for a depressed individual is mostly below the borderline as most of the features are taking a lower value than the normal.
4.2. Evaluation of the Depression Detection Model

The goal of this research is to detect depression through the analysis of selected user comments. The initial step is to use the first level feature space of the dataset. Four main classifiers, each utilising all of the feature categories, are used to assess the relevance of the various features obtained from the dataset. Evaluation metrics are used to analyse the performance of the aforementioned techniques. In this study, the following evaluation metrics are taken into account: i) Accuracy, ii) Precision, iii) Recall, and iv) F-measure. These values are based on information from the confusion matrix, which

Figure 10. Mental health curve of user from depressed dataset

![Figure 10](image1.png)

Figure 11. Mental health curve of user from normal dataset

![Figure 11](image2.png)
includes the expected result for each test sample. The experiments are evaluated utilizing 10-fold cross validation on all testing data levels.

Using classifiers, we evaluated the first-level feature performance and discovered improved prediction. An SVM classifier combined with all of the first-level features produces an accuracy of 84 percent and an F1-score of 0.77. Following this is the SA+PT+SMF classification model with LR (83 percent 0.76).

Sentiment analysis of a user’s profile and tweets is a crucial characteristic for figuring out their mental state, as demonstrated by the fact that when used as a single feature, Sentiment Analysis (SA) has an accuracy of 80% and an F1-Score of .79. The combination of Sentiment and Personality characteristics (ST+PT) surpasses the single SA feature (83 percent, .78) because knowing the user’s personality gives critical insight into the user’s behaviour on social media.

Temporal analysis (TA) gives useful information since the value of prediction is Yes in more than 40% of scenarios when the time is in AM. When used as a single feature, TA has an accuracy of 79% and an F1-score of 74% using XGBoost Classifier. The emoticons that a user uses to post material are examined to determine the Emotional Analysis (EA) score, which is a feature that assesses the user’s emotional state at the time. EA is therefore a crucial factor in determining how to interpret users’ mental states. Even while EA alone can’t achieve an acceptable level of accuracy (75 percent, 0.71), it excels when used in conjunction with other classifiers.

The social media feature (SMF) has an accuracy of 79% as a single level feature, indicating that we may use user profile information as an essential indicator for understanding the individual.

It obtains an accuracy of 83 percent and an F1 score of 0.72 when paired with additional significant features like PT and SA.

We used classifiers to assess how well the second level linguistic features performed and discovered that prediction was enhanced. Trigram+TF-IDF+LDA, a linguistic feature combination, outperforms feature combination with trigrams (87 percent, .87). For bigram and trigram features, as well as when the features are combined with LDA, SVM performs better than the other classifiers. In order to determine which combination works the best in terms of accuracy for the categorization of depressed tweets, we have now combined the first and second level variables. We utilised the Linguistic feature to supplement the Sentiment Analysis feature. In our investigation, integrating all of the characteristics (Trigram+SA+PT+EA+TA) resulted in the greatest performance for depression detection. It exceeded other feature combinations, including LDA+Trigram+TF-IDF+SA+EA (88 percent, 0.87), LDA+Unigram+TF-IDF+SA+EA+PT+SMF (86 percent, 0.85), LDA+Bigram+TF-IDF+PT+TA+SMF (85 percent, 0.85), and Bigram+TF-IDF+, with an accuracy of 89 percent and an F1-score of 0.88 with SVM (77 percent, 0.75). In most cases, the SVM is the best classifier, followed by Ensemble techniques, LR and XGBoost. A thorough study of the outcomes for combined attributes is shown in Table 5. Figure 12 compares the effectiveness of different classifiers.

5. DISCUSSIONS

Social media analysis is frequently used to generate insights for raising productivity and performance across a variety of applications. Social media is becoming more diverse, thus it’s important to completely understand the approaches and trends. Numerous studies indicate that Twitter is the most widely used social media platform for analysis. This analysis could take a while to complete manually. As a result, several analytical techniques are used, with classification and regression appearing to be the most popular.

There are many symptoms and indicators of depression, including the inability to concentrate on any task, to do homework, or to take pleasure in any activity. People who are depressed have feelings of disappointment, irritability, worry, and they think they are unfortunate and at fault. They experience fatigue, illness, headaches, and depressed thoughts. These parameters led to the identification of a number of indications and symptoms in the dataset, as well as the extraction of features related
to emotional variables (sad, positive, negative, anxiety, despair, and rage), language features, and
temporal features. Numerous feature extraction techniques, including sentiment analysis, emoticon
analysis, temporal analysis, and personality trait extraction, were used to extract these components.
Each tweet is examined sequentially, line by line to uncover hidden features.

The research proposes a novel method for calculating a multimodal mental health index for each
user using data collected from online media. Each feature has a range of normalised numerical scores,
and users are given values for the features based on whether their language contains any indication
of depression.

<table>
<thead>
<tr>
<th></th>
<th>LR</th>
<th>RF</th>
<th>SVM</th>
<th>XGBoost</th>
<th>Ensemble</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unigram+TF-IDF+SA+EA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.71</td>
<td>0.73</td>
<td>0.77</td>
<td>0.71</td>
<td>0.76</td>
</tr>
<tr>
<td>Precision</td>
<td>0.72</td>
<td>0.72</td>
<td>0.75</td>
<td>0.72</td>
<td>0.73</td>
</tr>
<tr>
<td>Recall</td>
<td>0.70</td>
<td>0.70</td>
<td>0.73</td>
<td>0.70</td>
<td>0.71</td>
</tr>
<tr>
<td>F1-Score</td>
<td>0.71</td>
<td>0.71</td>
<td>0.75</td>
<td>0.71</td>
<td>0.73</td>
</tr>
<tr>
<td>Bigram+TF-IDF+SA+EA+PT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.82</td>
<td>0.65</td>
<td>0.84</td>
<td>0.80</td>
<td>0.85</td>
</tr>
<tr>
<td>Precision</td>
<td>0.81</td>
<td>0.72</td>
<td>0.83</td>
<td>0.79</td>
<td>0.82</td>
</tr>
<tr>
<td>Recall</td>
<td>0.82</td>
<td>0.56</td>
<td>0.83</td>
<td>0.79</td>
<td>0.80</td>
</tr>
<tr>
<td>F1-Score</td>
<td>0.81</td>
<td>0.52</td>
<td>0.83</td>
<td>0.79</td>
<td>0.80</td>
</tr>
<tr>
<td>Trigram+TF-IDF+SA+EA+PT+TA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.84</td>
<td>0.63</td>
<td>0.89</td>
<td>0.81</td>
<td>0.89</td>
</tr>
<tr>
<td>Precision</td>
<td>0.83</td>
<td>0.80</td>
<td>0.88</td>
<td>0.80</td>
<td>0.87</td>
</tr>
<tr>
<td>Recall</td>
<td>0.84</td>
<td>0.53</td>
<td>0.87</td>
<td>0.80</td>
<td>0.88</td>
</tr>
<tr>
<td>F1-Score</td>
<td>0.84</td>
<td>0.44</td>
<td>0.88</td>
<td>0.80</td>
<td>0.88</td>
</tr>
<tr>
<td>LDA+Unigram+TF-IDF+SA+EA+PT+SMF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.86</td>
<td>0.61</td>
<td>0.70</td>
<td>0.71</td>
<td>0.87</td>
</tr>
<tr>
<td>Precision</td>
<td>0.85</td>
<td>0.30</td>
<td>0.70</td>
<td>0.69</td>
<td>0.84</td>
</tr>
<tr>
<td>Recall</td>
<td>0.85</td>
<td>0.50</td>
<td>0.70</td>
<td>0.69</td>
<td>0.86</td>
</tr>
<tr>
<td>F1-Score</td>
<td>0.85</td>
<td>0.38</td>
<td>0.70</td>
<td>0.71</td>
<td>0.86</td>
</tr>
<tr>
<td>LDA+Bigram+TF-IDF+PT+TA+SMF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.85</td>
<td>0.61</td>
<td>0.87</td>
<td>0.81</td>
<td>0.87</td>
</tr>
<tr>
<td>Precision</td>
<td>0.84</td>
<td>0.81</td>
<td>0.86</td>
<td>0.80</td>
<td>0.86</td>
</tr>
<tr>
<td>Recall</td>
<td>0.84</td>
<td>0.50</td>
<td>0.87</td>
<td>0.79</td>
<td>0.87</td>
</tr>
<tr>
<td>F1-Score</td>
<td>0.85</td>
<td>0.39</td>
<td>0.87</td>
<td>0.80</td>
<td>0.87</td>
</tr>
<tr>
<td>LDA+Trigram+TF-IDF+SA+EA+PT+TA+SMF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.84</td>
<td>0.63</td>
<td>0.89</td>
<td>0.81</td>
<td>0.89</td>
</tr>
<tr>
<td>Precision</td>
<td>0.83</td>
<td>0.80</td>
<td>0.88</td>
<td>0.80</td>
<td>0.83</td>
</tr>
<tr>
<td>Recall</td>
<td>0.84</td>
<td>0.53</td>
<td>0.87</td>
<td>0.80</td>
<td>0.78</td>
</tr>
<tr>
<td>F1-Score</td>
<td>0.84</td>
<td>0.44</td>
<td>0.88</td>
<td>0.80</td>
<td>0.85</td>
</tr>
</tbody>
</table>
Several studies involving sentiment analysis, emoticon analysis, temporal analysis, and tweet statistics are conducted to find a relationship between attitude toward depression and the feature used. The relationship between the aforementioned features and the attitude toward depression is explored throughout the text.

We applied multiple machine learning methods to assess the execution achieved by the extracted features. The goal was to find the most accurate combination of linguistic, topical, statistical, chronological, and emotional factors for the classification of depressed thoughts in tweets. Sentiment analysis and temporal variables were combined with linguistic qualities. From the data, it is clear that the SVM classifier performs optimally when all of the characteristics are combined. The Logistic Regression classifier has the second-best performance.

6. CONCLUSION AND FUTURE WORK

In order to better understand the public perception of depression, this study uses the classifier algorithms to identify emotional depressive phrases in user tweets. The authors carried out experiments to show that all of the aspects, including Linguistic, Topic, Temporal, Sentiment, and Emotional factors, are useful for identifying depression in tweets. When using the XGBoost classifier, the first level features—which are the results of looking at user profiles and tweet data—perform well. Using SVM and XGBoost classifiers, the first level metrics perform well when paired with language features. According to the results, it can be inferred that studying user profile data offers useful insight into the personality attributes of the user. The semantic meaning of user postings may be understood using the second level linguistic and topic features. A high degree of prediction performance may be reached by carefully choosing features.

The significance of the assessment metrics shows that there is still room for improvement and research despite the methodology’s effectiveness. Despite the fact that this study looks at temporal data, we have simply divided the day into four time periods (12PM-6AM, 6AM-12PM, 12PM-6PM, 6PM-12AM). We want to increase the precision of identifying depressive thoughts by using new feature sets, such as time information and word embedding. In order to predict individuals’ personalities based
on their social media activity, a thorough examination of personality and sentiment measurements may be helpful.

The mental health curve generated by using the multimodal features uses only the user’s online activities for calculating the mental health index. The curve would be able to provide more accurate prediction if other physical aspects such as heartbeat, blood pressure etc. can be incorporated for continuous multimodal surveillance of a person’s mental health. In order to understand more about the pattern of the user posting time and how it ties with depressive thoughts, the day can be divided into more time intervals. The current research provides a novel method as the groundwork for future research on the real-time surveillance of person mental health, discovery of new knowledge, such as the identification of the causes of depression and suicidal ideation and their potential ramifications. More feature sets, such as image information, time information etc. might improve the detection of depressive and suicidal thoughts. This research is expected to serve as a basis for future work in this area.
REFERENCES


Moumita Chatterjee currently works as an Assistant Professor in the Department of Computer Science and Engineering, Aliah University, Newtown, Kolkata. She received an M.Sc degree in Computer Science and an M.Tech degree in Computer Science and Engineering from the University of Calcutta, Kolkata. She was awarded the UGC NET-JRF fellowship in Computer Science. She is pursuing her Ph.D. in the domain of Distributed Computing from the same university. She is the author of several publications in refereed national/international journals and conferences, including IEEE, Springer, Elsevier, and many others. She has served as a reviewer of many international journals and conferences including IEEE Journal of Biomedical and Health Informatics, Journal of Supercomputing, ICECCME2023, OTCON 2023, and many more conferences. She has more than 5 years of experience in teaching and research. Her current research interests include Distributed Computing, Natural Language Processing, Machine Learning and Big Data technologies.

Piyush Kumar is currently working with Accenture Services Pvt. Ltd., Kolkata as a Advanced App Engineering Associate. He has completed his MCA from Amity University Kolkata. His research interests include social network analysis, data mining, artificial intelligence, and automation.

Dhrubasish Sarkar is currently working with Supreme Knowledge Foundation, Mankundu, Hooghly, West Bengal as a Professor of Computer Science and Information Technology at Supreme Institute of Management & Technology. He has more than 26 years of experience in teaching, training, research, academic administration, and institutional development. He holds PGCACS, MCA, M.Tech (IT) and PhD (CSE). He has more than 40 publications including reputed international journals, conference proceedings and book chapters to his credit. Currently he is serving as editorial review board member for few reputed international journals. His research interest includes Computational Social Science, Recommendation System, Machine Learning etc. He is a professional member of ACM.