Detection of Depression Trends in Literary Cyber Writers Using Sentiment Analysis and Machine Learning

Faiza Nasir[1], Haseeb Ahmad[1], CM Nadeem Faisal[1], Qaisar Abbas[2], Mubarak Albathan[2,*], Ayyaz Hussain[3]

[1] Department of Computer Science, University of Engineering and Technology, Taxila, Pakistan
[2] Department of Computer Science, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh 11432, Saudi Arabia
[3] Department of Computer Science, Quaid-i-Azam University, Islamabad, 44000, Pakistan

Summary
Rice is an important food crop for most of the population in Nowadays, psychologists consider social media an important tool to examine mental disorders. Among these disorders, depression is one of the most common yet least cured disease Since abundant of writers having extensive followers express their feelings on social media and de-pression is significantly increasing, thus, exploring the literary text shared on social media may provide multidimensional features of depressive behaviors: (1) Background: Several studies observed that depressive data contains certain language styles and self-expressing pronouns, but current study provides the evidence that posts appearing with self-expressing pronouns and depressive language styles contain high emotional temperatures. Therefore, the main objective of this study is to examine the literary cyber writers' posts for discovering the symptomatic signs of depression. For this purpose, our research emphases on extracting the data from writers' public social media pages, blogs, and communities; (3) Results: To examine the emotional temperatures and sentences usage between depressive and not depressive groups, we employed the SentiStrength algorithm as a psycholinguistic method, TF-IDF and N-Gram for ranked phrases extraction, and Latent Dirichlet Allocation for topic modelling of the extracted phrases. The results unearth the strong connection between depression and negative emotional temperatures in writer's posts. Moreover, we used Naïve Bayes, Support Vector Machines, Random Forest, and Decision Tree algorithms to validate the classification of depressive and not depressive in terms of sentences, phrases and topics. The results reveal that comparing with others, Support Vectors Machines algorithm validates the classification while attaining highest 79% f-score; (4) Conclusions: Experimental results show that the proposed system outperformed for detection of de-pression trends in literary cyber writers using sentiment analysis.

Manuscript received March 5, 2023
Manuscript revised March 20, 2023
https://doi.org/10.22937/IJCSNS.2023.23.3.7

Key words: *Depression detection; SentiStrength; TF-IDF; N-Gram; Classification; Machine Learning; Sentiment Analysis*

1. Introduction

World Health Organization, around 264 million people of all ages suffer from the de-pression worldwide. Suicidal deaths because of severe depression are increasing especially in young people. Such extreme acts may be avoided if such depression conditions may be detected and subsequently treated on time. Depression and anxiety are even visible in how a person speaks and writes [2]. More precisely, TM and NLP applications make it is possible to determine the depression-related contents from the social media posts that are publicly appeared [3]. The researchers have tried to reveal the association between the use of language and mental health conditions [4, 5, 6]. Such studies can be very useful to examine the individual's behaviors for predicting and subsequently controlling the destructive incidents. The researchers have discovered that the persons suffering from depression or anxiety are more prone to express negative emotions in text than the people of control (not depressive) group [5].

Few studies have observed behavioral and morphological attributes connected with depression and suicide [6]. The researchers have tried to analyze the poetic work through the corpus of suicidal notes available publicly to study suicidal behaviors. The authors used the popular psycholinguistic Linguistic Inquiry and Word Count (LIWC) lexicon to recognize the linguistic traces revealed by suicidal poets in their poems. The findings reveal that depressed poets use more first-person singular nouns than not depressive poets. Overall suicidal and depressive thoughts are linked with obstruction from the social interests and higher connection with self. Researchers also found that depressive poets include clue about death in their writings [7].

In many studies, a psycholinguistic lexicon-based LIWC analysis revealed that linguistic styles can be used to analyze the depression [6-10]. Secondly, topics have also been used to classify the contents [10-12]. In precise terms, the topic modeling may enable the understanding of the topics and language styles that depressive people usually use; hence, these methods provide the solution to differentiate between ab-normal health issues and the others. In addition, studies performed sentiment analysis alone and combined with other TM techniques to study depression among users [13-16]. Some researchers combined the LIWC psycholinguistic analysis tool with N-Gram and topic modeling techniques to study the language use of depressive individuals [9, 10]. Another study used word frequency-based clustering to differentiate the depressive and not depressive user dictionaries [17].

Several studies highlighted the depression on social media. But very limited among them discussed the depression in current trends related to cyber literary writers. In the prior studies, the adopted datasets and methods are repeatedly used for analysis while incorporating same features. There is a need to identify additional features and more deeper insights to overcome the real challenges. Although, both N-Gram and topic modeling have been used together in many studies, however, N-Gram, topic modeling along with emotional temperature measurement approach can even reveal better in-sights regarding depressive behaviors. Through this, contents of followees (followed by the followers) may also be classified between depressive and not depressive, subsequently, appropriate recommendations can be made for the followers.

Cyber-writers in our study are the people who use electronic medium and especially the online social media to express their emotional response, opinions, or expressions. In this age of social media, many people are sharing their poems, stories, blogs, personal writings on different social media platforms. These write-ups are called cyber-literature [18].

Nowadays, cyber literature has become the latest trend to be followed. Authors and writers have their public fan pages and communities where they share their thoughts. The emotions and language shared by the writers may also indicate the feelings of guilt, worthlessness, self-hatred, and helplessness, which are the symptoms of depression [19]. This "language of depression" affects the followers who read and follow their de-pressive contents.

This study intends to explore the literary user's posts to detect depressive features. Besides, we aim to examine the emotional temperatures and emotional terms used by these writers to classify depressive and not depressive behaviors. In precise terms, the proposal reveals deeper insights in four folds. First, we analyze emotional temperatures, then we extract specific depression-related keywords that are passed to the SentiStrength to check the emotional temperature values. Secondly, we analyze N-Gram phrases extracted through TF-IDF to identify the emotional terms language usage. Thirdly, we use LDA topic modelling to analyze topics from the extracted phrases/language. Subsequently, we employ four machine learning classifiers including Naïve Bayes (NB), Support Vector Machines (SVM), Random Forest (RF) and Decision Tree (DT) to validate the classification based on selected features and their combinations.

1.1. Major Contributions

It is important to consider that the performance metrics primarily depend upon selected features along with their combinations [9]. Our findings show that writers with depression express more stress-related emotions than relaxation. More precisely, the writers express the negative traits as compared to not depressive. Our main contributions are enlisted as follows:

- To crawl cyber-writer's data for classification of the depressive and not depressive posts.
- To present a feature selection approach based on emotional temperatures, keyword-based approach (the emotional temperatures around the specific keywords including personal pronouns, self-expressing pronouns, and depression-related words).
- To reveal linguistic topics through LDA of depressive phrases extracted through N-Gram and TF-IDF.
- To validate the classification results based on single feature and a combi-nation of features by incorporating machine learning classifiers.

1.2. Paper Organization

The remaining contents of the paper are organized as follows. Section 2 briefs about the literature review. The detailed methodology is presented in Section 3. Section 4 compiles the results and discussion. Section 5 concludes the paper and spots light on future work.

2. Literature Review

Psychologists believe that depressive people behave differently as compared to not depressive people. These differences may be successfully identified through their social media activities, language style, status, word choices, and writing styles that they normally use to write the posts [20]. Prevalent studies have discussed such behavioural issues and uncovered several depression-related features. These studies revealed a strong relationship between depression and the use of negative emotions. Moreover, the usage of first-person, second-person pronouns, and negative thoughts related to life, self-harm attitudes, regrets, and disappointments are also revealed [21].

Research has been focused to determine the association between the use of language and mental health conditions. Technologies are really supporting us for reaching out to these aspects in precise manners. Linguistic analysis has been performed to ex-tract insights in this field [22]. Few studies used sentiment analysis methods to classify a given text into a class depending on the given time and reason. Another streamline is the classification that can have different types including binary (positive or negative), trinary, multi-class, or dual [23-25]. Such classification may help to extract linguistic features that humans are unable to recognize efficiently, for instance, calculating the frequencies of words and grammatical structures of the text [26]. In parallel, few studies discussed the depression detection through social media contents [27-29].

The bloggers community is considered as well-known among online communities [30]. The huge number of contents in the form of blogs and related comments are available to the researchers for developing the comprehensive algorithms for clustering of information [31]. The imperative focus of the researchers lies in the contextual meanings of opinions and sentiments [32], community sense [33, 34], political views [35] and user interactivity concerns [36].

Topic modelling has been extensively used for revealing the intended topics of the written contents. Working towards this domain, Yin et al. [11] used the Latent Dirichlet Allocation to classify topics between different groups and the employed algorithm produced significant results. Hence, TM and NLP are the latest trends to automatically extract deeper insights. Comments, keywords, and opinions provide important indicators about the author's behavior through a multidimensional investigation of un-structured data [37]. Techniques for grouping the data generally includes clustering, sentiment analysis and data mining [14, 37]. Using the aforementioned tools, the re-searchers have tried to analyze the poetic work through the corpus of suicidal notes available publicly to study the suicidal behaviors [7]. However, fewer studies have observed behavioural and morphological attributes connected with depression and suicide. The need of the hour is to find the effective features to classify the contents.

2.1 Language of depressive writers

Understanding of the language contents and style specifications used by depressive people in their posts depict that how the depressive individuals think. Combining automated text analysis with machine learning may effectively classify mental health issues from samples including tweets, Reddit, blog posts [38]. More, precisely the con-text of writings can be differentiated using the contents and the style of expressing words. The meaning of statements related to what we express normally lies in the contents. It can easily be identified that depressive users use more stress-related negative emotional features, especially negative adverbs, and adjectives, for example, "gloom", "lonely", "lost", "tired", "sad", "miserable". Moreover, the depressive users use more self-expressing pronouns like I, my, me, mine, myself etc. [20]. Therefore, such words may be used as features to identify depression symptoms in cyber-writers (see Figure 1).

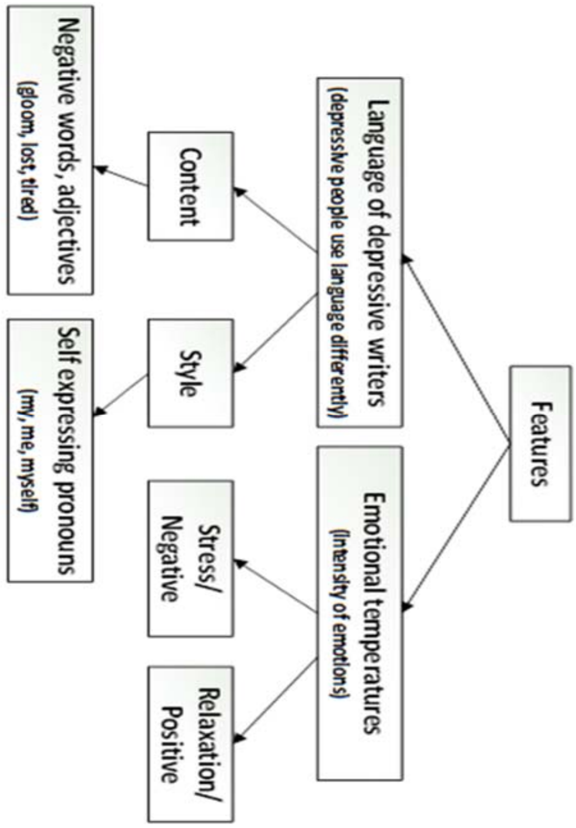


Figure 1. Features to classify depressive and not depressive Users.

2.2. Emotional Temperatures

The emotional temperature is the intensity of emotions in the text [39]. The emotional intensities depend upon the situation or circumstances in which the person is expressing a sentiment. The two main categories of emotions we consider are defined as follows:

2.2.1 Stress or Negative Feelings

Stress or negative feelings are usually produced by good and bad experiences [24]. On social media, people suffering from different mood swings use the same language and words differently, which identifies their current mood at specific time.

2.2.2. Relaxation or Positive Feeling

Relaxation in psychology is the emotional state of a living being [24]. Relaxation is the positive state free from tension, anxiety, and other types of stress. In the relaxation state, the individuals generally use positive words in their text that depict positive attitude. Although, the words emotion and sentiment have a slightly different meaning in text analysis, however, still these are used synonymously [39]. The human expressions are divided into two perceivable measurements of intensity/temperature and polarity. The polarity labels the human expression in positive, negative, or neutral sentiment while the intensity measures the strength of the sentiment or emotion and is reported in a numeric value [40]. Both the polarity and temperature values totally depend upon the contents and the style of the language used in the text to be analysed. For example, if a text contains more negative words, adjectives, adverbs, intensifiers or negation with positive words, the emotional temperature will be scored towards negative polarity.

Getting intuition from such advancements, in this work, we are identifying the depressive symptoms in literary pieces of cyber writers. For classification, we examine the emotional polarities and word usage patterns between the depressive and not de-pressive groups. Moreover, we are incorporating the distinctive features to classify the depressive and not depressive contents, and subsequently validate the classification through machine learning classifiers.

Table 1: Description of raw data

Data	Description
Sample size	6547
Sample type	Blog posts, personal writings, writer's pages, quotes, poems, etc.
Platform	Facebook (3860), blogs posts (2687)
Words	100,435
Tokens	130456

3. Proposed Methodology

3.1. Data Acquisition

For this research, the benchmark dataset was not available; therefore, we scraped the data from different cyber writer communities and pages through publicly available social media posts. We explored hashtags for love, life, depression, anxiety, hope, loneliness, art, poetry, literature, writers, writing with depression to collect the literary data. Our data consist of a total of 6547 posts from Facebook and blogs. Table 1 depicts the description of raw data used in this research.

3.2 Proposed Methodology

In this study, we employ a hybrid approach detailed as follows:

1) Emotional temperatures are extracted through SentiStrength
2) N-Gram features are extracted through TF-IDF for depression identification.
3) LDA is used for topic modelling.

The research model starts from data collection and pre-processing followed by feature extraction by using the three approaches. Subsequently, the language/phrases classification is validated through four state-of-the-art machine learning classifiers. The results are visualized and discussed as shown in Figure 2.

3.2.1 Proposed Algorithm

At this stage, we applied the SentiStrength algorithm to study whether the emotional temperatures can be used to classify the depressive and not depressive writers. SentiStrength is currently the powerful lexical based sentiment analysis algorithm that can classify negative, positive, or mixed sentiments [24, 25, 29, 39]. It calculates the stress and relaxation scores based on the polarity of terms, intensifiers, exclamation marks, booster words, and negations. Furthermore, it stores the emotional terms, unigram, bigram, and N-Gram feature vectors which are then used to classify the text. Sensi Strength's original dictionary is domain-independent and partially derived from LIWC, one of the most popular psycholinguistic lexicons. But it has been developed to classify common user text, routine chats, reviews, etc. [29]. To acquire the domain-specific results extend the default lexicon with a domain-specific lexicon. Therefore, to achieve the accurate depressive text classifications, we adapted SentiStrength by extending its lex-icon with our domain-specific words validated by domain experts (University Professors from Psychology and English). The model in Figure 4 illustrates the step-by-step procedure of lexical extension process and classification.

The most essential part of sentiment analysis is to apprehend the context-specific use of words [40]. Therefore, the domain-specific dictionary extension is very important to achieve the target accuracy. Our dataset does not entirely comprise literature or poetry, but every part of write-up that was shared on writer's pages and blogs. Hence, we do not focus only on poetic literature terms, but also the words associated with de-pression. To collect and add these words and phrases, we explored the dataset, de-pression-related studies, and online articles, and included all depression, anxiety, suicidal words in the lexicon after validating these by domain experts. The SentiStrength's default lexicon typical score scale starts from 1 (no emotion) to 5 (high level) for positive emotions and -1 (absence of (stress) to -5 (strong presence of stress) for negative emotions but these scores can be changed to obtain the domain-specific accuracy [24, 25]. There are two ways; annotation of the text by independent human coders or by using the default lexicon to annotate the text, then increasing or decreasing these scores according to the type of the analysis [24, 25]. We adopted the second one. We made a list including all the words in our dataset. Subsequently, we classified all texts by a SentiStrength algorithm using the default lexicon.

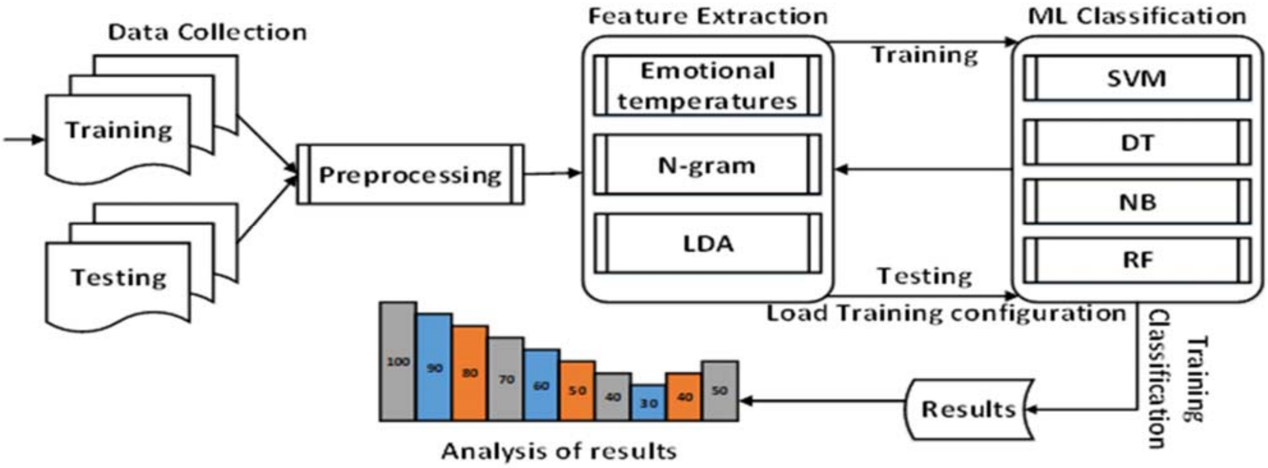


Figure 2. A systematic flow diagram of proposed model for classification of depression patterns.

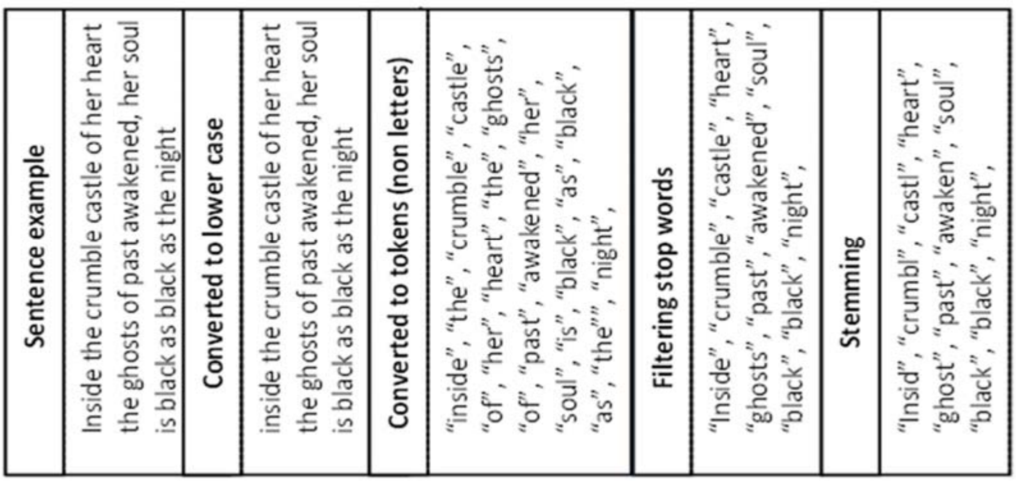


Figure 3. A sample of pre-processing words of stemming.

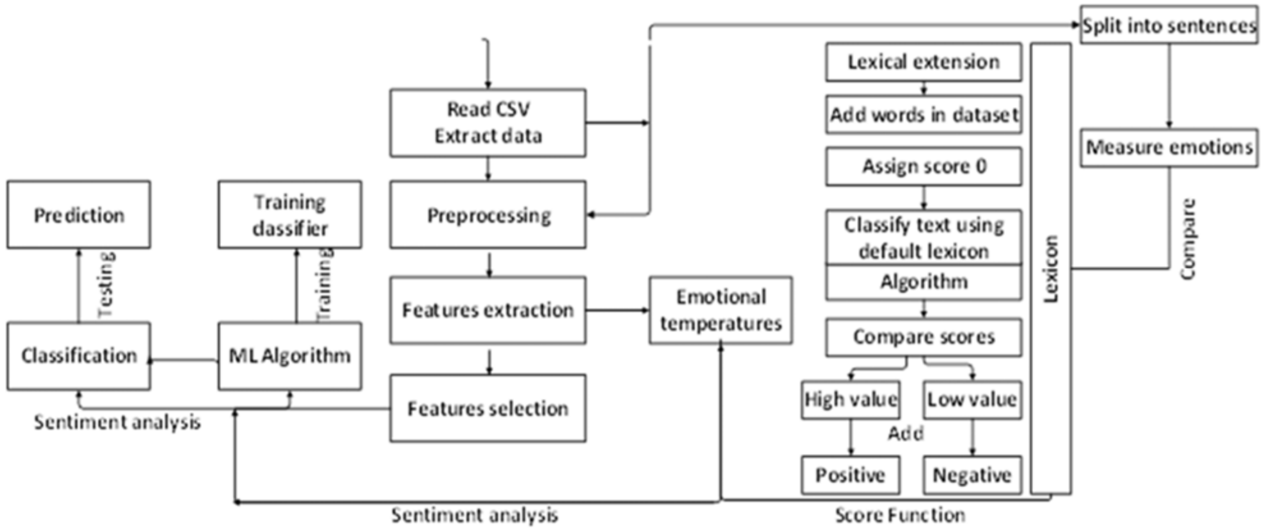


Figure 4. Measurement of emotional temperatures using sentiment analysis and score function.

Table 2: Lexicon construction words with respect to score.

Word	Score
Loneliness	-4
Depressed	-5
Frustrated	-3
Gloomy	-4
Grumpy	-4
Guilty	-4
Irritated	-3
Melancholy	-4
Rejected	-3
Sadness	-4
Hopelessness	-4
Emptiness	-3
Tiredness	-3
Insomnia	-4
Anxiety	-5
Worthlessness	-4
Suicidal thoughts	-5

The algorithm assigns scores on general basis without considering the domain specificity, but these scores can be analyzed and altered to enhance the domain specific classification. The method is to initially assign 0 score to all the text. Afterwards, analyzing the text through SentiStrength algorithm and comparing these scores with de-fault lexicon. Then, increasing or decreasing the scores of domain specific words ac-cording to the analysis domain as illustrated in Figure 4. Therefore, we analyzed the initial classification and increased the scores of the depression-related words by 1 to differentiate the mood between depressive and not depressive writers. For example, the original algorithm assigns different strength scores to depression and anxiety, but we used -5 for these domain-specific keywords to make it work on depression recognition domain. The scored words are then added into the lexicon after validating them from domain experts. Some of these words and their scores are shown in Table 2.

3.2.2 Measurement of Emotional Temperatures

Initially, the dataset was unlabeled i.e., without any knowledge of depression among writers with only writer ids in the user column and their mined posts in a text column. To divide the dataset into depressive and not depressive indicative text samples, we first classified it through SentiStrength algorithm's binary classification mode. Binary classification mode assigns positive or negative polarity to each text. Then, we manually assigned class labels to both the groups based on the examined polarities (positive or negative). Later, we analyzed the emotional temperatures in two ways. Initially, we analyzed the stress and relaxation intensities for each entry in the dataset. Then, we applied the keyword matching approach to analyze the emotional intensities relevant to that specific keyword in both groups. Keyword matching is used to analyze the emotional temperatures around the specific keyword. We fed the depression indicative keywords and personal pronouns to the algorithm to analyze the emotional temperatures in the posts containing these keywords.

As the dataset contains one or two-line quotes and few long paragraphs, i.e., articles, blog posts. To calculate the overall stress and relaxation, we divided the paragraphs P into separate sentences denoted as P = (s1, s2, …, sm) during the pre-processing phase. Subsequently, for each sentence si, stress and relaxation values are calculated to measure the overall stress and relaxation values of the whole post. The total of all emotions in all lines indicates the overall sentiments of the article or paragraph [39]. The higher stress values as compared to relaxation ascertain the presence of depressive mood. The emotional temperatures are based on the mood and intensity of emotions in writing [39]. The higher polarity of negative emotions over positive depicts the higher negative emotional temperatures, use of negative sentiments that are the key elements in de-pression detection.

Moreover, our dataset shows significant imbalance class distribution. This class imbalance in the training dataset made the classification algorithms biased towards the majority class and we got 95% accuracy. The classifier which is biased is useless to provide correct prediction. To overcome this problem, we used Synthetic Minority Oversampling Technique (SMOTE) filter [43]. SMOTE filter generates similar data samples for minority class samples and gives balanced data samples in both classes. We set the nearest neighbour parameter value to 5 and percentage value to 100.

To evaluate the prediction power of emotional temperatures, classification with four classifiers is performed. For this purpose, the data was stored in CSV format file containing the first column for user-id, the second column for the text to be classified, third column for depressive or not depressive annotation for each of the text rows, fourth and fifth columns were containing stress and relaxation values, respectively. We used percentage split 66% to segregate the data into training and testing. The correctly classified text in the dataset is crucial for accurate predictions. More precisely, the words with higher stress and lower relaxation value were labelled as depressive, while that with higher or equal relaxation value comparative to stress were labelled as not de-pressive. Neutral words were considered in the not depressive group. The algorithm can misinterpret results in some cases such as in the absence of direct emotions in a sentence, or in absence of proper negation in critical sentences, we manually observed the sentences that were generated neutral and changed the sentiment to positive or negative accordingly, subsequently validated it by domain experts. For example, the sentence "I need a wish to stop wishing" was scored as neutral sentiment because it doesn't contain any direct emotion term, so we had to change it manually because it depicts depressive meaning.

3.2.3 Features Extraction

The N-Gram is a sequence of n words. N-Gram is widely used in NLP and TM as a method to calculate the likelihood of co-occurrences of each sentence in a dataset as a uni-gram, bi-gram, tri-gram or N-Gram [9]. We computed TF-IDF to generate N-Grams feature vector. TF-IDF is the term inverse document frequency, which is calculated to recognize the importance of a word for document in the collection of documents. It has been extensively used in text classification studies to extract valuable features and is still very popular [20, 42]. We selected a total of 50 uni-gram, bi-gram, and trigram phrases to which IDF assigned higher weights among all. We set the maximum length of N-Gram to 3. The N-Grams were computed for both depressive and not depressive classes (see example in Table 3).

Table 3: The examples of the most frequent N-Gram features.

N-Gram in the depression group	N-Gram in the not depressive group
more~pain~more	something~new
breath~to~endure,	see~dream
more~to~bear	new~journey
turn~into~ashes	embracing~my~passions
fragment~of~myself	be~free
my~soul~bang	Love~lies
useless~for~you	with~my~coffee
it~also~ache	i'm~happier
being~alone	felt~butterflies
heart~suffer	cherish~the~moments
not~worth	to~happy
to~depart	you~share
not~aliv	Forget~
I~lie	
stop~wish	
hide~behind~smile	
myself~alone	
almost~dead	

Latent Dirichlet allocation (LDA) probabilistic models are widely used in topic modelling to discover topic structures from text documents [9-12]. Each document in LDA is a probability allocation over the latent topic and each topic is a distribution over terms in the group of documents. The LDA model summarizes the features by clustering similar words into topics. It is important to specify the number of topics. Each topic is clustered with certain words. The number of words are also specified before topic generation. LDA finds similarity in topics and assigns specific weights to each topic in documents and to each word in the topic. We used LDA topic modelling to extract and analyze the topics in depressive and not depressive groups. Figure 5 illustrates the LDA modelling process.

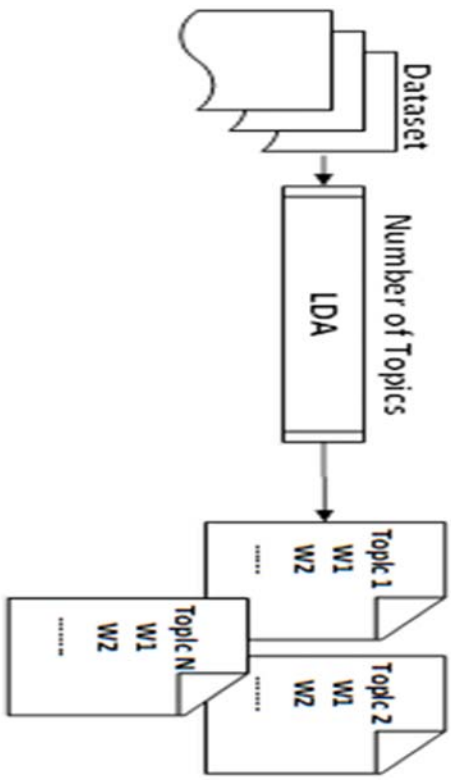


Figure 5. Topic modelling using LDA approach.

To analyze and select the most active topic, we executed the LDA model multiple times with 1000 iterations to select the optimal topics. After getting lower accuracy results, we increased the number of topics by 10 iteratively, obtaining the best performance with 100 topics containing not more than 12 words (with maximum weights) in each topic. We took a higher number of topics because the performance of LDA gets better as more topics are added [41]. We selected the topics containing the maximum distance in both classes with not more than 10 words (with maximum weights) in each topic during iterations.

4. Results and Discussion

In this research, we studied the depressive symptoms in literary pieces of cyber writers. To classify the depressive writers, we examined the emotional polarities and word usage patterns between the depressive and not depressive groups.

4.1 Emotional temperatures analysis

The initial psycholinguistic classification results of the original dataset 6537 (see Table 4) reveal the stress and relaxation means and standard deviation in depressive and not depressive samples. Among these, a total of 4445 (68%) are found depressive containing the higher stress and lower relaxation mean values and 2091 (31%) as not de-pressive with the higher relaxation and lower stress scores. Although, direct mentions of depression word count are only 15% of the sample. The depressive class has higher stress levels as compared to relaxation. The mean value of stress identified in depressive class is (-3.97) on a scale of -1 not negative to -5 very negative, whereas the relaxation mean value is only (1.38) on a scale of 1 not positive to 5 very positive. The mean score of stress in not depressive class is (1.77), which is less than the stress scores of depressive classes. Similarly, relaxation has a higher mean value equal to (2.80) in not depressive class while the depressive class has a relaxation mean value 1.38 only. Furthermore, the stress levels are almost identical in the writers who used personal pronouns in their writing, but the relaxation element is only (0.57) in depressive people. According to the keyword analysis results, most of the users who used personal pronouns are classified in the depressive category. Although, sentiments including wishes, reminiscence, love, life, and feel were discussed in both groups, however, the stress levels are relatively high in depressive writers, which means the depressive group shared these emotions in a more negative way.

The relaxation levels in not depressive users are higher than depressive users, which mean that depressive users talk about negative aspects of life, as they express negative feelings. Their write-ups contain sadness, worthlessness, suicidal thoughts, and death-related words. They talk about their worthlessness; they are more likely to use past tenses, dreams, lost time and unfulfilled or fulfilled wishes. The results of emotional temperature analysis around the specific keyword and its synonyms (in brackets) are also shown in the Table 4. The stress level in not depressive users while expressing the life aspects is very low (1.9) and it is (3.01) in depression class text samples.

About death, stress and relaxation both are high in depressive users that mean the depressive users may find relief in death or escape. Sadness elements are clearly high towards negative scores (3.4) in depressive writers and relaxation is only (1). Anger and worthlessness expressions are also high with (1.9) towards negative emotions. In joy, related sentiments, stress levels are slightly different with (1.56) in depressive and (1.73) in not depressive writers, but relaxation levels are higher (1.77) in not depressive class and (1.04) in depressive. Alongside, while talking about love, depressed users showed disappointment and negativity because the stress elements are higher (2.7) than not depressive writers (1.9). Stress levels in remembering past events (1.88) and using past tenses (2.94) are higher in depressive writers which mean that depressive people used past tenses in their writings, and they expressed negative emotions while they wrote about their memories. There is no such notable difference found in the usage of the second person pronoun and first-person plural, although first-person singular pronouns are mostly found in higher negatively expressed write-ups with only (0.5) relaxations.

4.2 N-Gram analysis

To examine depressive content, we analysed the dataset and selected the top 50 N-Grams in depressive and not depressive group. We first computed the term frequency and then inverse document frequency of the words. We selected the words with higher IDF weights. The higher size words as compared to the rest of the words indicate a high-frequency usage of words in depressive class. The words frequency analysis of depressive and not depressive posts in Figure 6 indicates the mood differences between two classes. The frequencies of words time and past are extremely high in depressive writers' posts. The choice of words sleeps, die, cry, self, black, end and scream characterize unhappy mood and depressive thoughts of writers.

The not depressive writers' posts are found to be the quite opposite as these reveal their happy and neutral mood, happiness, excitement, and overall positive or neutral emotions.

Figure 6. N-Gram in the depressive and not depressive group.

4.3 Topic analysis using LDA

We built LDA topic model to extract the hidden topics between depressive and not depressive groups. The model clusters the topic with words in depressive and not depressive posts [10]. To understand the visualization, it is important to consider the LDA measures that each document contains specific part of a certain topic. The closest topics are more similar, the distant topics are less similar, and each topic represents a collection of most representative words. The greater the size of the word, the more common they are in the collection of documents [11].

The optimal classification is achieved with perplexity 30, 1000 iterations and 100 topics with words containing higher weights. As we explained earlier, we executed the model multiple times by changing these values to get the reported accuracy.

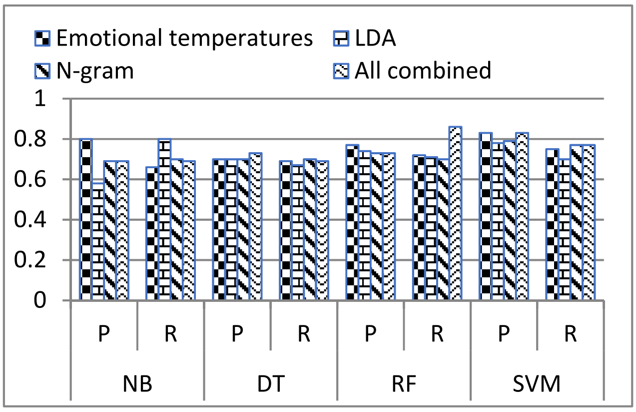
Figures 7-10 represent the topics and words correlation in each topic. Figures 7 and 8 represent the topics in depressive group. In topic 1, the words cry, pain, suffer and alone are strong depressive predictors. In topic 2, words feel, think, love, dream, wish and wait are more frequent. It signifies that the writers, who shared their feelings or talked about love, also shared their unsatisfied dreams, and wishes. In topic 3, most prominent words are time and past which represent that the writers are likely to remember past times, their distress and use of drug. In addition, the writers shared haplessness, life failures and suicidal thoughts in topic 4.

However, the topic clusters of normal (not depressive) writers in Figures 9 and 10 are predictors of normal behavior. The word clusters walk, rain, breathe, calm and lovely in topic 1 relates to nature's pleasure. Topic 2 also have positive and neutral words i.e., success, music, live and money. The words life and hope are present in both classes' topic clusters, but the other words in cluster around these words indicate the usage difference. Topic 3 and 4 clusters show the writers talk about help, motivation, possibilities, their social meetups, friends, family support, and beauty of life. They shared helping and supporting emotions, their family events, and happy times with friends, respectively.

4.4 Results Analysis

In classification, the first step is to calculate a baseline metric and subsequently to compare the results of other classifiers with it. Baseline classifier calculates the most frequent class. The notion here is to build a model that gives better predictions than the baseline results. For this purpose, we use Zero R classifier as a baseline for data classification problems. Zero rule is the naive classifier which flouts all predictors and depend on the target. It is used to predict the baseline accuracy and predicts the most frequent class instances in training dataset. The predictions are used to evaluate all other classifiers performance. If the model generates better results than baseline classifier, then it is considered as a better model. The results are shown in Figure 11. We got the Precision (0.45), Recall (1.0) and F-score (0.64) while classifying through Zero R.

To create a depression detection model, we used two datasets, labelled in the training phase and unlabeled in the testing phase. The predicted labels identify the presence of depression among the writers. We trained and tested the model on all features one by one and then combined results were obtained. We applied four major classifiers including NB, SVM, RF and DT to classify the depressive and not depressive samples. For the implementation of these classifiers, we used 10-fold cross-validation with 66-34% split. To evaluate the classification performance, we calculated the precision Eq. (1), recall Eq. (2) and F-score Eq. (3). Precision defines the number of positive



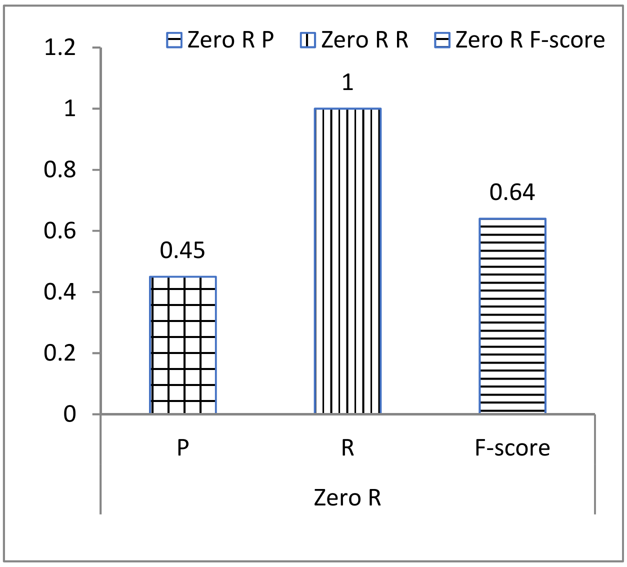


Figure 11. Baseline Precision, Recall, F-score of Zero R classifier

Figure 12. Precision and Recall scores of all features and classifiers.

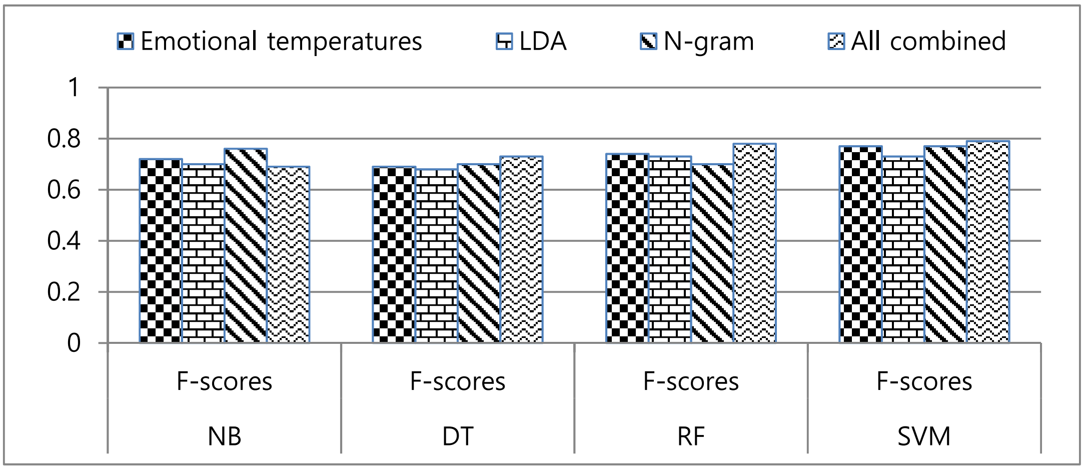


Figure 13. Comparison of all feature and classifiers performance.

Table 4: Mean and standard deviation scores (max value can be 5) of psycholinguistic analysis for the de-pressive and not depressive users' posts.

Selected Features	**Depressive class (4445)**		**Not depressive class (2091)**	
Emotional Temperatures	Stress	Relaxation	Stress	Relaxation
Overall (6547)	3.97(0.77)	1.38(0.78)	1.77(0.85)	2.80(0.7)
First person pronoun (I, me, myself)	1.98(0.62)	0.57(0.77)	1.93(0.77)	1.99(0.5)
First person plural (We, us, our)	1.90.(0.62)	1.93(0.6)	1.90(0.62)	1.94(0.7)
Second person plurals (you, your, yours)	1.80(0.6)	1.29(0.6)	1.90(0.65)	1.4(0.75)
Past (was, had, been, did)	2.94(0.7)	1.66(0.68)	1.7(0.68)	1.7(0.75)
Memory, reminiscence	1.88(0.7)	1.30(0.76)	1.24(0.6)	1.35(0.6)
Wish, desire, longing, hope, crave, dream	1.94(0.9)	2.3(0.7)	1.90(0.7)	1.83(0.4)
Love, Liking, Desire, passion, attraction, lust	2.7(0.7)	2.56(0.5)	1.96(0.66)	1.77(0.61)
Joy, happiness, bliss, ecstasy, cheer, satisfy	1.56(0.54)	1.04(0.6)	1.73(0.66)	1.77(0.61)
Worthless	1.98(0.76)	1.03(0.78)	1.0(0.7)	1.0(0.5)
Anger, Hate, Disgust, Annoyance, Dislike, Bitterness	1.98(0.62)	1.57(0.67)	1.03(0.8)	1.79(0.5)
Sadness, Hurt, Gloom, Agony, Depression, Anxiety, Death, Loneliness, Hopelessness, Melancholy, Guilt, Shame, Regret, Reject, Insult, Moan, Sad, Lost, Alone, Pain	3.40(0.26)	1.05(0.05)	1.90(0.62)	1.94(0.7)
Death, Suicide, Bury, Kill, Depart, End	2.0(0.6)	1.38(0.07)	1.66(0.85)	1.0(0.05)
Feel	2.38(0.62)	1.57(0.09)	1.30(0.85)	1.99(0.5)
Life, Alive, Life, Living, survival, Journey	3.01(1.9)	1.98(1.1)	1.20(0.12)	2.5(0.67)

(depressive) predictions that truly belong to positive group. Precision enumerates the validity of predictions. Recall is the number of positive predictions from total positive instances in the data. Recall tells the completeness of the prediction results. F1- score is the harmonic mean which is calculated from recall and precision of the test. These values are calculated by the following formulas.

In binary classification problem and imbalance dataset, F-score represents the measure of correctness. The F-score is more important than simple accuracy measures. It determines the correct classification ability of the classifier. Its value varies from 0 to 1. The value 1 indicates the optimal classification performance of the classifier and the 0 indicates the worst classification, which indicates the poor precision and recall. The Figure 12 illustrates the results of precision, recall and Figure 13 represents F-score of each classifier and predictive power of features and classifier's performance. The high F-scores (0.79) are achieved by SVM classifier with all combined features. The high precision (0.83) signifies the classifier's ability to predict correct depressive instances. Furthermore, the emotional temperatures as a single feature obtained the second highest F-score (0.77) with SVM classifier. The classification metric for all other classifiers with n-gram and LDA is also reasonable. In addition, the high Recall is achieved by SVM classifier with all combined features. Overall, the combined features (LDA + N-Gram + Emotional Temperatures) produce the greatest classification correctness with 0.79 F-score.

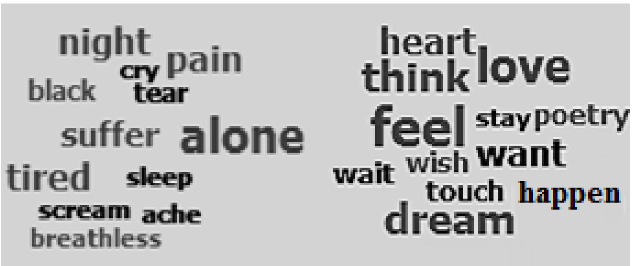


Figure 7. Topic distribution over words in Depressive group.

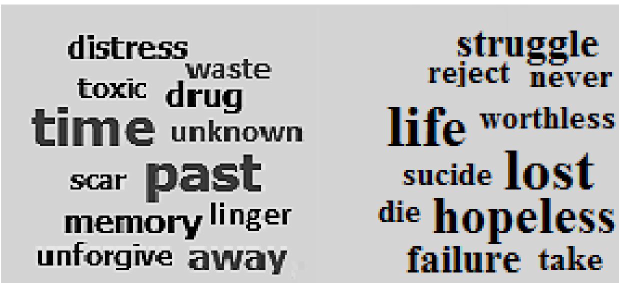


Figure 8. Topic distribution over words in depressive group.

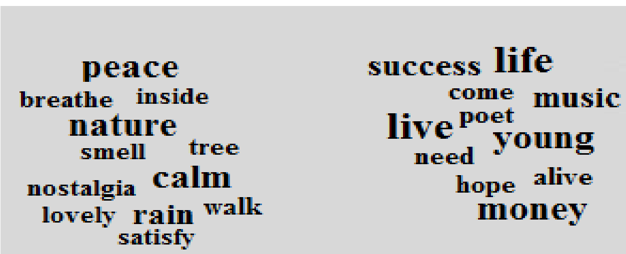


Figure 9. Topic distribution over words in not depressive group.

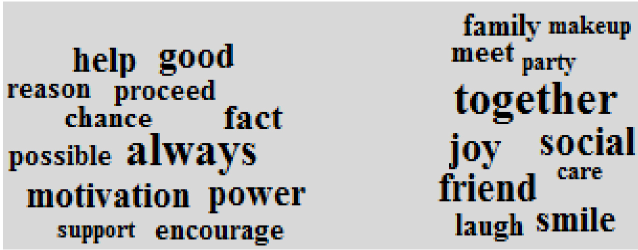


Figure 10. Topic distribution over words in not depressive group

5. Conclusion

In this research, we studied the depressive symptoms in literary pieces of cyber writers. To classify the depressive writers, we examined the emotional polarities and word usage patterns between the depressive and not depressive groups. We used SentiStrength algorithm as a psycholinguistic method, TF-IDF and N-Gram for ranked phrases extraction, and Latent Dirichlet Allocation for topic modelling of the extracted phrases. We examined the emotional temperatures and mixed emotions including stress and relaxation levels in each post. Moreover, we applied sentiment keyword technique to observe the usage of personal, impersonal pronouns, expression of emotions along with usage of death and suicide-related words and feelings in contents of depressive and not depressive writers. The mean values in each case of our findings indicate that depressive writers use death-related words and negative aspects of life. The results further reveal that depressive writers express negative emotions towards past events, lost love and unfulfilled wishes. Alongside, such writers express a hopeless attitude in their posts. While the not depressive writers talk about the topics expressed satisfaction, joy and positivity. Therefore, their relaxation levels were higher than depressive users. The study observed that most of the writers on social media and blogs are depressive while only 2091 samples are found in the positive class. The results show that higher stress intensities, negative polarities are great indicators of depression. The N-Gram and LDA extracted phrases characterize the language usage between depressive and not depressive writers. The language styles of the depressive user are different in that they use negative emotions, expressed their memories, worthless feelings and negative sentiments about life, relationships, love, and survival in their write-ups. According to the classification results, the emotional temperatures are the best predictors of depression among all single features with SVM precision (0.83), recall (0.75), f-score (0.77). In addition, all combined features with SVM classifier are the best predictors of depression with precision (0.83), recall (0.77), f-score (0.79). The study reports significant insights; however, few limitations are present that need to be improved in the future. Firstly, the data sample is small, which can prevent many other important features to be observed. Secondly, the lexicon construction and depression words annotation mechanism require improvements to enhance the validity. Moreover, this study is applied only on users and their write-ups; there are other important attributes that can give more effective results.

In the future, our focus is to study other important dimensions such as followers, likes and locations that would make more effective classifications. Followers and followed persons are notable elements as they can be used to study the mood trends of connected users, which may also expose more depressive survivors.

Acknowledgments

The authors extend their appreciation to the Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University (IMSIU) for funding and supporting this work through Research Partnership Program no. RP-21-07-11.

References

[1] A. G. Reece and C. M. Danforth, "Instagram photos reveal predictive markers of depression," EPJ Data Science, vol. 6, p. 15, 2017.
[2] World Health Organization, "Depression and other common mental disorders: global health estimates," World Heal. Organ., pp. 1–24, 2017.
[3] P. Howlader and A. Cuzzocrea. "Predicting Facebook-Users' Personality based on Status and Linguistic Features via Flexible Regression Analysis Techniques," In Proceedings of the 33rd Annual ACM Symposium on Applied Computing (SAC '18), pp. 339–345, 2018.
[4] C. Tai, Z. Tan, and Y. Lin, "Mental Disorder Detection and Measurement using Latent Dirichlet Allocation and SentiWordNet," IEEE International Conference on Systems, Man, and Cybernetics, pp. 1215–1220, Oct. 2015.
[5] A. Abboute, Y. Boudjeriou, G. Entringer, J. Azé, S. Bringay, and P. Poncelet, "Mining twitter for suicide prevention," in International Conference on Applications of Natural Language to Data Bases/Information Systems, pp. 250-253, 2014.
[6] A. Jeri-Yabar et al., "Association between social media use (Twitter, Instagram, Facebook) and depressive symptoms: Are Twitter users at higher risk?", International Journal of Social Psychiatry, vol. 65, no. 1, pp. 14-19, 2018.
[7] W. Count, "Word Use in the Poetry of Suicidal and Nonsuicidal Poets," Psychosomatic Medicine, vol. 522, no. 12, pp. 517–522, 2001.
[8] T. Nguyen, D. Phung, B. Dao, S. Venkatesh, and M. Berk, ''Affective and content analysis of online depression communities,'' IEEE Trans. Affect. Comput., vol. 5, no. 3, pp. 217–226, Jul. 2014.
[9] M. M. Tadesse, H. Lin, B. O. Xu, and L. Yang, "Detection of Depression-Related Posts in Reddit Social Media Forum," IEEE Access, vol. 3536, no. c, pp. 1–12, 2019.
[10] J. Wolohan, M. Hiraga, A. Mukherjee, Z. A. Sayyed, and M. Millard, "Detecting Linguistic Traces of Depression in Topic-Restricted Text: Attending to Self-Stigmatized Depression with NLP," in Proceedings of the First International Workshop on Language Cognition and Computational Models, pp. 11-21, 2018.
[11] Z. Yin, L. Cao, Q. Gu and J. Han, "Latent Community Topic Analysis", ACM Transactions on Intelligent Systems and Technology, vol. 3, no. 4, pp. 1-21, 2012. Available: 10.1145/2337542.2337548.
[12] W. Speier, M. K. Ong, and C. W. Arnold, "Using phrases and document metadata to improve topic modeling of clinical reports," Journal of Biomedical Informatics, vol. 61, pp. 260–266, 2016.
[13] C. Lin, E. Ibeke, A. Wyner and F. Guerin, "Sentiment-topic modeling in text mining", Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 5, no. 5, pp. 246-254, 2015. Available: 10.1002/widm.1161.
[14] X. Wang, C. Zhang, Y. Ji, L. Sun, L. Wu and Z. Bao, "A Depression Detection Model Based on Sentiment Analysis in Micro-blog Social Network", Lecture Notes in Computer Science, pp. 201-213, 2013.
[15] C. H. E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text," in Eighth International Conference on Weblogs and Social Media (ICWSM-14). Available at (20/04/16) http://comp. social. gatech. edu/papers/icwsm14. vader. hutto. pdf, 2014.
[16] J. Neidhardt, N. Rümmele, and H. Werthner, "Predicting happiness: user interactions and sentiment analysis in an online travel forum," Information Technology & Tourism, vol. 17, pp. 101-119, 2017.
[17] H. Park, M. Lee, S. Hwang and S. Oh, "TF-IDF Based Association Rule Analysis System for Medical Data," KIPS Transactions on Software and Data Engineering, vol. 5, no. 3, pp. 145-154, 2016. Available: 10.3745/ktsde.2016.5.3.145.
[18] Rahman, F. "Cyber Literature: A Reader – Writer Interactivity," International Journal of Social Sciences & Educational Studies, vol. 3, no. 4, pp. 156-164, 2017.
[19] M. De Choudhury, S. Counts, E. J. Horvitz, and A. Hoff, "Characterizing and predicting postpartum depression from shared facebook data," in Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing, pp. 626-638, 2014.
[20] M. De Choudhury, M. Gamon, S. Counts, and E. Horvitz, "Predicting depression via social media," ICWSM, vol. 13, pp. 1-10, 2013.
[21] A. Abboute, Y. Boudjeriou, G. Entringer, J. Azé, S. Bringay, and P. Poncelet, "Mining twitter for suicide prevention," in International Conference on Applications of Natural Language to Data Bases/Information Systems, pp. 250-253, 2014.
[22] M. Trotzek, S. Koitka, and C. M. Friedrich, "Utilizing Neural Networks and Linguistic Metadata for Early Detection of Depression Indications in Text Sequences," pp. 1–14, 2018.
[23] Wang, M.; Shi, H. Research on sentiment analysis technology and polarity computation of sentiment words. In Proceedings of the 2010 IEEE International Conference on

Progress in Informatics and Computing, Shanghai, China, 10–12 December 2010; pp. 331–334, 2010.

[24] M. Thelwall, "TensiStrength: Stress and relaxation magnitude detection for social media texts," Information Processing & Management, vol. 53, pp. 106-121, 2017.

[25] M. Thelwall, "Heart and soul: Sentiment strength detection in the social web with sentistrength, 2013," Cyberemotions: Collective emotions in cyberspace, 2013.

[26] Y. R. Tausczik and J. W. Pennebaker. “The psychological meaning of words: LIWC and computerized text analysis methods,” Journal of Language and Social Psychology, 29(1):24–54, 2010.

[27] Mikal J, Hurst S, Conway M. Investigating patient attitudes towards the use of social media data to augment depression diagnosis and treatment: a qualitative study. In: Proceedings of the fourth workshop on computational linguistics and clinical psychology—from linguistic signal to clinical reality. 2017.

[28] J. Abbas et al., "The moderating role of social support for marital adjustment, depression, anxiety, and stress: Evidence from Pakistani working and nonworking women", Journal of Affective Disorders, vol. 244, pp. 231-238, 2019.

[29] Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., Kappas, A.: Sentiment strength detection in short informal text. Journal of the American Society for Information Science and Technology, 61(12), 2444–2558, 2010.

[30] Z. LIU, Y. FU, X. LING and X. XIONG, "Blog community detection based on formal concept analysis", Journal of Computer Applications, vol. 33, no. 1, pp. 189-191, 2013.

[31] R. Yesha and A. Gangopadhyay, "A method for analyzing health behavior in online forums," in Proceedings of the 6th ACM Conference on Bioinformatics, Computational Biology and Health Informatics, pp. 615-621, 2015.

[32] A. Park and M. Conway, "Longitudinal changes in psychological states in online health community members: understanding the long-term effects of participating in an online depression community," Journal of medical Internet research, vol. 19, 2017.

[33] Z. LIU, Y. FU, X. LING and X. XIONG, "Blog community detection based on formal concept analysis", Journal of Computer Applications, vol. 33, no. 1, pp. 189-191, 2013.

[34] G. Nimrod, "From knowledge to hope: Online depression communities," International Journal on Disability and Human Development, vol. 11, pp. 23-30, 2012.

[35] Laganas C, McLeod K, Lowe E. “Political posts on Facebook: an examination of voting, perceived intelligence, and motivations,” Pepperdine J Commun Res. Vol. 5 no. 1, p.18, 2017.

[36] J. A. K. Johnsen, J. H. Rosenvinge, and D. Gammon, "Online group interaction and mental health: An analysis of three online discussion forums," Scandinavian journal of psychology, vol. 43, pp. 445-449, 2012.

[37] Islam MR, Kamal ARM, Sultana N, Moni MA, Islam R. Depression detection using K-Nearest Neighbors (KNN) classification technique. in International conference on computer, communication, chemical, materials and electronic engineering, February 8–9, 2018, Rajshahi, Bangladesh.

[38] M. Islam, M. Kabir, A. Ahmed, A. Kamal, H. Wang and A. Ulhaq, "Depression detection from social network data using machine learning techniques", Health Information Science and Systems, vol. 6, no. 1, 2018. Available: 10.1007/s13755-018-0046-0.

[39] J. Culpeper, A. Findlay, B. Cortese, and M. Thelwall, "Measuring emotional temperatures in Shakespeare’s drama," English Text Construction, vol. 11, pp. 10-37, 2018.

[40] R. Islam and M. F. Zibran, “SentiStrength-SE: Exploiting domain specificity for improved sentiment analysis in software engineering text,” The Journal of Systems & Software, vol. 145, no. August, pp. 125–146, 2018.

[41] Y. Tyshchenko, "Depression and anxiety detection from blog posts data", 2018.

[42] M. Hiraga, "Predicting Depression for Japanese Blog Text", in Proceedings of ACL 2017, Student Research Workshop, Vancouver, Canada, 2017, pp. 107–113.

[43] Chawla, Nitesh V., Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. "SMOTE: synthetic minority over-sampling technique." Journal of artificial intelligence research 16 (2002): 321-357.