UNIVERSITY OF BIRMINGHAM

Research at Birmingham

Analysis of user-generated content from online social communities to characterize and predict depression degree

Fatima, Iram; Mukhtar, Hamid; Ahmad, Hafiz Farooq ; Rajpoot, Kashif

[Link to publication on Research at Birmingham portal](#)

Analysis of User-Generated Content from Online Social Communities to Characterize and Predict Depression Degree

Iram Fatima
College of Computer Sciences and Information Technology, King Faisal University, Saudi Arabia

Hamid Mukhtar
School of Electrical Engineering and Computer Science, National University of Sciences and Technology, Pakistan

Hafiz Farooq Ahmad
College of Computer Sciences and Information Technology, King Faisal University, Saudi Arabia

Kashif Rajpoot
School of Electrical Engineering and Computer Science, National University of Sciences and Technology, Pakistan
School of Computer Science, University of Birmingham, UK

Abstract
The identification of a mental disorder at its early stages is a challenging task because it requires clinical interventions that may not be feasible in many cases. Social media such as online communities and blog posts have shown some promising features to help detect and characterize mental disorder at an early stage. In this work, we make use of user-generated content to identify depression and further characterize its degree of severity. We used the user-generated post contents and its associated mood tag to understand and differentiate the linguistic style and sentiments of the user content. We applied machine learning and statistical analysis methods to discriminate the depressive posts and communities from non-depressive ones. The depression degree of a depressed post is identified by using variations of valence values based on the mood tag. The proposed methodology achieved 90%, 95% and 92% accuracy for the classification of depressive posts, depressive communities and depression degree, respectively.

Keywords
Depression classification; mental health; moods and emotions; online communities; depression degree identification

1. Introduction

Social media tools such as blogs and online discussion forums have become increasingly recognized as open and free communication platforms to help in problem solving and information sharing. Recently, there has been a growing research interest in the use of social media for identification, prevention, or intervention of different kinds of mental illnesses [1] [2]. Due to recent lifestyle changes, every human being undergoes the feelings of tension, anxiety, or sadness at different times. When these feelings become so disturbing and overwhelming that people have great difficulty in coping with the day-to-day activities [3][4] such as work, enjoying leisure time, and maintaining relationships then it is considered an indication of some mental illness. Medical, psychological and social experts have identified that there are more than 450 different definitions and types of mental illnesses with varying degrees of severity[1]. Mental illnesses are estimated to account for 11% to 27% of the disability burden in Europe [3], while mental disorders are the leading cause of years lived with disability worldwide [4]. Some of the major types of mental illnesses are depression, anxiety, bipolar disorder, personality disorder and schizophrenia. Of these, the most common mental illnesses are anxiety and depressive disorders[2]. The World Health Organization (WHO) has ranked major depression, a common form of mental illness, as one of the most burdensome diseases in the world [5].

There is a need for effective interventions, policies, and prevention strategies to allow early detection and diagnosis of mental health concerns in populations. Traditionally, most of the assessment is carried out using questionnaires requiring a subjective response or comment by the patient. Generally, such responses are not only influenced by the context – the

environment and the patient's relationship with the clinician – they may also be a representation of the patient's state of mind, or mood, at that particular instance in time and not the actual, prevailing state of mind of the patient [2]. Social media, on the other hand, not only allows its users to express their thoughts in their own words but at a time when they feel the need to express. Social support from social media is crucial to well-being and quality of life of patients with incurable and recurrable diseases [6]. A series of thoughts and expressions over a longer period provides a better opportunity for the assessment process. Therefore, social media provides a rich source of author-identified text that can be used for personality profiling as well as knowing the mental state of a person. For this reason, social media has been recognized as an important tool [7] [8] for identifying and analyzing depression. In general, different text mining techniques are used for the analysis of user's social media posts. For evaluating mental health conditions, researchers use different textual cues including writing style, word usage, sentence structure, vocabulary, topic of the text, etc. In particular, for identification of depression, researchers have found the use of swear words and expression of sadness as a feature of the text [9].

A relatively unexplored territory in the analysis of depression from the text is the use of mood and emotions. Efforts such as [10] [11] have analyzed mood and emotion in the text but did not evaluate it for identification of depressive symptoms. Most of these efforts depend upon a dictionary of words related to various degrees of moods – called affective lexicon – for mood identification. Perhaps, a reason for not associating emotions with depression is that, despite its simplicity, it has been found that creation of an effective lexicon is difficult since only 4% of words used in texts have emotional value [12].

In this study, we demonstrate that by using only a small subset of language features, not only we can differentiate between depressive and non-depressive text but we can also identify the degree of depression with high accuracy. Although much of the recent research work in depression analysis has used topic modelling, we are not using this approach because of two reasons. First, use of topics restricts the analysis to a handful of topics as it is not possible to take care of all the topics expressed in all types of blogs. A topic based approach will typically result in an adequate analysis in a particular set of topics for which the algorithms are trained but it will result in relatively poor analysis when applied in the wild. Second, topic analysis will falsely classify any posts as depressive that will only be informative in nature on the topic of depression, e.g., an article about depression written by some expert. Avoiding such misclassification is yet another challenge, which has not been addressed adequately thus far.

Our contributions in this work are two-fold. First, we use language style and sentiment information for finding the most effective data dimensions of Linguistic Inquiry Word-Count (LIWC) [12] by applying feature extraction. This approach provides a set of predictors for the classification of depressive posts [13] and communities. Second, together with the mood expressed by the author, we are also able to characterize the degree of the depression expressed in the text either as mild, moderate, or severe. To our knowledge, this is the first of its kind of work for analysis of depression degree after identification of the text as depressive text. Estimating the depression degree is important to determine the urgency and severity of treatment. In the literature, Hamilton Depression Rating Scale [14] is a popular tool that classifies depression into four degrees: mild, moderate, severe and very severe. For this research, we are using the first three only as the identification of a person having a severe degree of depression is as important to treat urgently as someone with very severe condition. To allow other researchers to build upon our work, we explain the different machine learning and statistical methods used in our analysis in great detail along with results and a discussion on them.

The rest of the paper is organized as follows. We briefly discuss the background and related work in Section 2. In Section 3, we introduce the proposed methodology for the classification of depressive posts and communities followed by the prediction of depression degree. In Section 4, we analyze and evaluate the experimental results to validate the proposed approach. Finally, the conclusion is presented in Section 5.

2. Background and Related Work

Social media platforms have become a rich source of information about individuals for recording their individual-centric thoughts, feelings, or opinions about small and big happenings in their life [15]. The study in [16] highlighted the support of social media for creation of tacit knowledge and sharing. Their study found six main ways where social media can support information encountering and provided opportunities for users to gain greater value of knowledge creation and sharing. The advantage of measurement of behaviour via social media helps in capturing one's social activity and language expression in a naturalistic setting [17] [18] as compared to doing the same via traditional settings such as interviews, which typically require recollection of certain facts that might be subjective and may vary according to the participant's current mental state or mood [19].

The authors in [20] did a comparative analysis of Facebook, Twitter, Delicious, YouTube and Flicker to analyze the motivation of users for sharing information and social support in social media. They involved 1,056 social media users in

five different surveys about the motivation of sharing information to understand the human information behavior. The results showed that learning is the highly influential motivation and social engagement is the second. The study in [7] discovered a connection between social anhedonia and Twitter users. They collected the dataset by using Amazon crowdsourcing. They conducted the depression survey followed by questions related to depression history and demographics with the access to the Twitter accounts of the participants. Their results indicated that a depressed person typically has a smaller social network, more negative feelings, greater concerns with drugs and intense expression of religious ideas. They developed a model using Support Vector Machines (SVM) classifier that predicts depression of an individual with 70% accuracy. The work done in [21] studies the impact of 14 words with the potential to stigmatize the mental health on Twitter. The data was collected in two stages (a) keyword based data and (b) user based data and their findings show that mental health aware users use stigmatizing words less frequently than other users. This indicates the sensitivity of users towards stigmatization of those with mental illnesses.

In a study conducted in 2011, Facebook profiles of 200 students were tested for the purpose of determining symptoms and depression level [22]. The findings of their research reported that 30 students' status updates show the indication of hopelessness, insomnia, or excessive sleeping. Their results concluded that the college students are facing more depression as compared to other people. The authors in [23] proposed an algorithm to detect stress and relaxation strength in tweets with a significant agreement rate with human judgements. They developed a lexical approach based system to detect the strength of stress and relaxation. The result showed that their proposed algorithm is flexible enough to work in a range of different contexts therefore; it can be used as an off-the-shelf solution for stress and relaxation detection. Park et al. [24] showed that online social network data can be successfully used for clinical studies. They performed sentiment analysis on tweets by using the LIWC [12]. They developed a multiple regression model by using all the sentiment categories and examined how variables of LIWC are associated with the CESD-R score [13].

In [25], the authors compared information need and provision by analyzing 10 depression blogs and 40 threads of Finnish internet discussion forums. They applied descriptive statistics and qualitative content analysis and identified that instead of factual and procedural information; most of the users were interested to get an opinion or evaluation of an issue relevant to depression. The work done in [26] analyse the use of affective information, topics and language style for depression community and personal blogs. Their results indicate that language style and topics have strong indicative powers for the prediction of depression. The authors in [27] discriminated online messages between depression and control communities using mood, psycholinguistic processes and content topics extracted from the posts generated by members of these communities. According to their research, writing style of both communities are significantly different that contributed in discriminating the depressive communities from control communities. Sentiment analysis shows the clinical group has lower valence than people in the control group. They extracted a number of features for affect, mood, linguistic style and topic of the post. For the affect feature, the Affective Norms for English Words (ANEW) lexicon [28] was used. To identify the mood of the user, they relied on the user-tagged mood label of the post. For linguistic style, the LIWC features [12] were used. Finally, they extracted topic for each post using the Latent Dirichlet Allocation (LDA) approach [29]. The major drawback of their approach is the application of a complex pipeline that involves a series of algorithms and may not be scalable to large datasets. Compared to them, our approach relies on a small number of features and does not carry out topic extraction, which is not only expensive but also restricted to only a few topics present in the dataset.

Malmasi et al. [30] carried out the classification of data from ReachOut.com forum posts into two main categories. The distinguishing feature of their work is that they employed a meta-classifier that used a set of base classifiers constructed from lexical, syntactic and metadata features. Initially, a single classifier was trained for each feature type and context, resulting in an ensemble of over 100 classifiers. The output from these classifiers was used to train a meta-classifier, which outperformed the individual classifiers as well as an ensemble classifier. This meta-classifier was then extended to random forests of meta-classifiers, yielding further improvements in classification accuracy. Although their classification achieved an overall accuracy of as high as 91% for categorizing a post into one of two labels – green and non-green posts – it was limited because of the nature of dataset: problems specific to youth population.

Saha et al. [31] developed a framework for classifying online mental-health-related communities for identification and presence of a mental condition such as depression. The framework used multi-task learning (MTL) as a joint learning method where an independent problem is considered as a task and MTL computes parameters of multiple tasks in an integrated framework. They used two main features of the text: language style and topic. The language style was extracted using the LIWC tool while topics from the posts were extracted using LDA [29]. A total of 68 topics and 50 linguistic features were used. Their suggested MTL framework outperformed a single-task approach. However, this work is limited due to the usage of topics and high-dimensional data for classification.

There is not much work in the area of mood classification for blog posts. Mishne [32] introduced one of the first mood classification methods from blog posts. They used the post length, word frequency, word's semantic orientation, emphasized words, and special symbols as features. The classification accuracy was modest, being slightly above baseline.

Nguyen et al. [11] used a wider range of features, including cheap and effective features inspired from psychology study, for the problem of mood classification for LiveJournal posts. The best accuracy result achieved was 78.8%. A better approach that used a hierarchy of possible moods was introduced in [10] [33], achieving better results than flat classification.

The existing work reviewed above has contributed significantly in finding the depression from user generated content posted on different social media. We first described studies that use Social Networking Sites (SNS) such as Facebook and Twitter for depression analysis [7][21-23]. Then we mentioned work done in the area of online blogs [11][26-32]. We also discussed the approaches for depression identification related to writing styles [24][27], use of lexicons [20][23][27], sentiment analysis [23][24][27], machine learning [23][30][31], and mood-based identification [10][11][32][33]. The focus of most of the research work was to identify and understand the differences in the writing style of the depressed individuals. However, in this paper, we build a model on existing findings with the additional feature of predicting the degree of depression as an important factor in determining the treatment urgency.

3. Methodology

The proposed framework consists of six major modules, as shown in Figure 1: (1) Data Extraction: to collect data from social communities for depression analysis. (2) Community Analysis: to apply LIWC to identify the variations of different sentiments in each community. (3) Feature Extraction: to identify the significant data dimensions to facilitate the classification algorithm for better performance. (4) Post and Community Classification: differentiate the depressive posts and communities from non-depressive posts and communities. (5) Depression Degree Analysis: analyze the depressive posts to measure the degree of depression and (6) Depression Degree Classification: assign a degree of depression to each depressive post. The details of each module are described in the following sections.
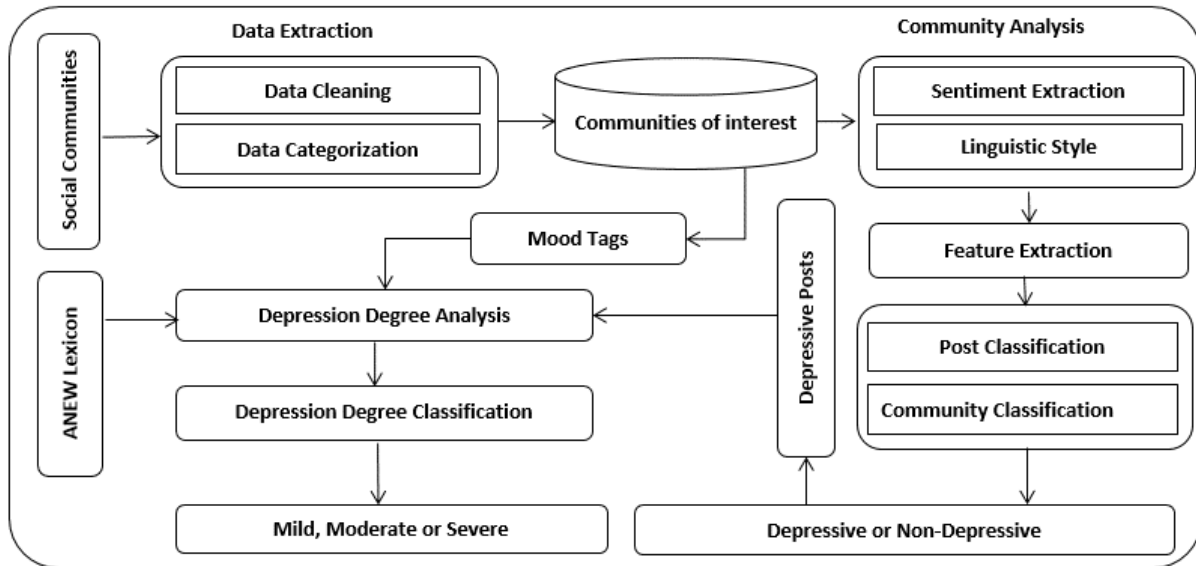

Figure 1. The proposed system architecture.

3.1. Data Extraction

Data is crawled from LiveJournal[3], a well-known platform for people to join their community of interest and discuss various issues. This most popular blogging site has attracted over 1.9 million active users since 1999 [34]. We identify

depressive and non-depressive communities by using "search communities by interest" option provided by LiveJournal. The depressive communities are selected based on the description of individual communities like depression, bi-polar, death, and suicide. The non-depressive communities are extracted by considering different aspects of life such as computer help, childcare, and beauty. After crawling the data, it was cleaned by removing unnecessary tags and labels and use the post title, post mood and post body for further data analysis.

3.2. Community Analysis

In different communities, people discuss various topics with positive and negative emotions. In depressed communities, people talk about health, anxiety, and sadness while in other communities the topics of discussion are home, jobs and leisure activities. In order to identify the difference between communities, LIWC features [35] are extracted to analyze the word use within text. LIWC calculates the percentage of usage of sets of words and assigns an output measure to different linguistic categories. Post title and post body are provided as the input and 93 output variables are produced by LIWC to indicate the variations in sentiments and linguistic style.

3.3. Feature Extraction

From LIWC results, we are interested in a set of variables that can help to differentiate the depressive posts and communities from rest of the data. For this purpose, we use RELIEFF [36][37] as a feature extractor that computes rank and weight of each data dimension by using regression with K-nearest neighbors. At each iteration, RELIEFF takes i^{th} feature vector y_i and computes its closeness to each class by Euclidean distance. The computed close class is called *Near* and the other is called *Far* as shown in equation 1.

$$W = W_i - (y_i - Near_i)^2 + (y_i - Far_i)^2 \qquad (1)$$

Table 1. Detailed explanation of the feature set.

Feature Set	Description	Example
I	first personal singular	I, me, mine
Posemo	positive emotion	love, nice, sweet
Negemo	negative emotion	hurt, ugly, nasty
Anx	Anxiety	worried, fearful, nervous
Cogproc	cognitive process	cause, know, ought
Insight	Insight	think, know, consider
Cause	Cause	because, effect, hence
Health	Health	clinic, flu, pill
Affiliation	Affiliation	ally, friend, social
Informal	informal language	shit, OK, hmm

RELIEFF assigns the weight vector based on nearby instances of the class. In our case, the input for the RELIEFF is 93 output variables from LIWC. RELIEFF computes the significance of 93 input variables by computing their closeness to depressive and non-depressive classes and as an output, we construct a feature vector consisting of top 10 variables based on the assigned weight and rank for each variable. The relevant weight and rank of remaining variables are significantly low and static in comparison to the top 10 variables. The detailed explanation of variables of the feature set is shown in Table 1. The first column represents the formal name of the feature attribute as given by LIWC, the second column shows its description and examples are given in the third column for clear understanding. The extracted feature set values serve as an input to the classification algorithm for the identification of depressive and non-depressive posts and communities.

3.4. Post and Community Classification

In this section, we examine the usefulness of extracted features by utilizing them in the two classification setups studied in this work. Firstly, posts are classified as depressive or non-depressive. Subsequently, set of posts (i.e. a community) is classified as depressive or non-depressive community. For this purpose, we used random forests [38] that is an ensemble learning method for classification, which operates by constructing decision trees at training time, and gives out the class label by using mode or mean of individual trees. Random forests use averaging of deep decision trees to reduce the variance by training on different parts of the same dataset.

The learning procedure of the random forests classifier starts by building the random trees and each tree casts a unit vote for the most popular class to classify an input vector. The design of decision tree requires the choice of attribute selection and a tree pruning method. In this work, the random forests classifier uses Gini index for attribute selection and measuring the relevance of an attribute with the class label. For a given training set X, selecting one class (depressive) at random and saying that it belongs to some class C_i, the Gini index [39] is shown in equation 2.

$$\sum\sum_{j\neq i}\left(\frac{f(C_i,X)}{|X|}\right)\left(\frac{f(C_j,X)}{|X|}\right) \tag{2}$$

where $f(Ci,X)/|X|$ is the probability that the selected case belongs to class C_i. One of the major benefits of the random forests classifier over the other decision tree methods is that trees those are grown to maximum depth on training data using combination of features are not pruned.

Algorithm: Post and Community Classification
Input: *Ts*-Train Set[]
             Tes-Test Set[]
             Cl-Class Label[]
             St=Post classification=0, Community Classification =1
Output: *Pc*-Predicted Class []
Begin
1   **for** i= 1: *length(Ts)*
2   │   *CleanData[i] = DataExtract (Ts [i])*
3   **end**
4   *AnalysedComm [] = LIWC (CleanData[])*
5   *[ranked , weight] = RELIEFF (AnalysedComm[])*
6   *FeatureSet [] = Top-10 ([ranked, weight])*
7   *nTrees = 50;*
8   *model = RandomForest (nTrees, FeatureSet[], Cl)*
9   **if** (*St* == 0)
10   │   *Pc = model.predict (Tes[])*
11   **else**
12   │   **for** *j = 1:length (Tes)*
13   │   *ClassfiedLabel[j] = model.predict (Tes[j])*
14   │   **end**
15   │   *Pc = MajVote (ClassfiedLabel[])*
16   **end**
End

Figure 2. Algorithm for Post and Community Classification.

The number of features to generate a tree and total number of trees to be grown are two user defined parameters required to generate a random forests classifier. In our case, we use feature set based on 10 attributes to generate a tree and we set

the total number of trees to 50. To classify a new feature set, each case is passed down to each of the 50 trees. The random forests classifier picks a class having the most votes for that class. The flow of whole classification process is shown in Figure 2.

3.5. Depression Degree Analysis

The content of each depressive post may differ and relevant mood tag helps us to identify the level of depression. Therefore, the focus of this section is to analyse the characteristics of depressed posts only and we use LiveJournal mood tags for this purpose. LiveJournal provides 132 pre-defined mood tags, thus providing a potential source to understand the affective aspect of a post. We categorize mood tags into three major categories: (a) severe depression, (b) moderate depression, and (c) mild depression. We use the ANEW lexicon [28] to map the mood tags to depression level. In this research, ANEW lexicon is used as a valid and useful tool that allows to manipulate the affective properties of different words and our focus is to explore the pre-defined list of LiveJournal under 1034 words of ANEW lexicon, rated in terms of valence and arousal.

We use the valence value for quantitative estimation of depression. Valence is a measure in psychology to categorize specific emotions. For example, the popular negative emotions such as anger and fear have low valences while positive emotions, events and situation such as joy and love have high valence values. The valence of ANEW words is on a scale of 1 (very unpleasant) to 9 (very pleasant). We use the scale of ANEW and set the range of valence for each level of depression: 1.0-3.5 (severe depression), 3.6-5.5 (moderate depression) and 5.6-9.0 (mild depression). The sample of moods categorization with their relevant ANEW valence values is shown in Table 2. This illustrates that moods which belong to severe depression have very low valence value in comparison to moods which represent moderate and mild depression.

Table 2. Moods categorization for depression degree analysis.

Severe Depression		Moderate Depression		Mild Depression	
Mood	Valence	Mood	Valence	Mood	Valence
Rejected	1.50	Scared	3.62	touched	6.31
Depressed	1.83	Listless	4.12	surprised	7.47
Frustrated	2.48	Lazy	4.38	thoughtful	7.65
aggravated	2.66	Indifferent	4.61	optimistic	7.59
Thirsty	3.37	Sympathetic	5.33	loved	8.72

3.6. Depression Degree Classification

In classification of depression degree, the objective is to label each post identified as depressive with the level of severity of depression found in the content. For the classification of depression degree into one of the three severity types described earlier using ANEW scale, we consider the assignment of depression degree as a classification problem. Not all the depressive posts have mood tags in LiveJournal communities so first, we set the degree of depression for the posts with mood tags and then we infer the depression degree of posts with missed mood tags. The process for identification of depression degree after analyzing the depression intensity is shown in Figure 3.

First, the algorithm is trained with given mood tags against three class labels: (a) severe depression, (b) moderate depression, and (c) mild depression by encoding them as 0, 1 and 2 respectively. For this purpose, Hierarchical Hidden Markov Model (HMM) [40] [41] is applied to the values of moods and class labels after converting them to discretized set. We choose HMM as it is a generative probabilistic graph model that is based on the Markov chains process and well known for labeling discrete sequences. The training model is based on the number of states (depression level) and their transition weight parameters. Parameters are learned through observation (mood tags) and the following parameters are required to train the model:

$$\lambda = \{A, B, \pi\} \tag{3}$$

where λ is a graphical model for depression level, A is a transition probability matrix, B represents the output symbol probability matrix, and π is the initial state probability [36]. We use Baum-Welch algorithm to determine the states and transition probabilities during the training of HMM. The ith classification weight of a post is given in equation 4.

$$\lambda_i = \{A_i, B_i, \pi_i\} \quad i = 1, \ldots, N \tag{4}$$

Where λ_i is the classification weight for i[th] class that belongs to one of the three class lables: (a) severe depression, (b) moderate depression, and (c) mild depression.

Algorithm : Depression Degree Analysis and Classification
Input: *Dp-* Depressive posts with mood tags[]
                 *Tes-*Test set without mood tags[]
Output: *Pdd-*Predicted degree of depression []
Begin
1   **for** i= 1: *length (Dp)*
2   │ *valence[i] = ANEW (Dp[i].mood-tag)*
3   │ **if** *(valence[i] >= 1 && valence[i] <= 3.5)*
4   │ │ *DepDeg[i] = 0   //severe-depression*
5   │ **else if** *(valence[i] > 3.5 && valence[i] <= 5.5)*
6   │ │ *DepDeg[i] = 1   //moderate-depression*
7   │ **else**
8   │ │ *DepDeg[i] = 2   //mild-depression*
9   │ ***end***
10 **end**
11   *model = HMM (Dp[], DepDeg[])*
12   *Pdd = model.predict (Tes)*
End

Figure 3. Algorithm for Depression Degree Analysis and Classification.

4. Results and Discussion

In this section, we present the results to evaluate and validate the feasibility of the proposed approach for classification of the depressive posts, depressive communities and assignment of depression degree.

4.1. Dataset Description

The experiments were performed on 10 communities from LiveJournal. We selected five depressive and five non-depressive communities as shown in Table 3. We obtained and analyzed a total of 4,026 posts, consisting of 2,019 depressive and 2,007 non-depressive posts. In Table 3, the 'Community Name' column shows official name of the community, '#Member' column shows the total number of members for a community in the dataset, '#posts' column represents the total number of posts we collected from each community for the sake of experiments and 'Description of the community' column shows general purpose of each community.

4.2. Performance Evaluation Measures

In order to evaluate the performance of the proposed system, the standard metrics of precision, recall, f-measure, and accuracy are used as performance evaluation measures. Their values are calculated using the confusion matrix [42] and computed as:

$$Precision = \frac{1}{Q}\sum_{i=1}^{Q}\frac{TP_i}{NI_i} \qquad (5)$$

$$Recall = \frac{1}{Q}\sum_{i=1}^{Q}\frac{TP_i}{NG_i} \qquad (6)$$

$$F - Measure = \frac{2 \cdot precision \cdot recall}{precision + recall} \qquad (7)$$

$$Accuracy = \frac{\frac{1}{Q}\sum_{i=1}^{Q}TP_i + \frac{1}{Q}\sum_{i=1}^{Q}TN_i}{Total} \qquad (8)$$

where Q is the number of posts, TP is the number of true positives, NI is the total number of inferred labels, TN is the total number of true negatives and NG is the total number of ground truth labels and $Total$ is the total number of depressive and non-depressive posts (or communities) in the dataset.

Table 3. Characteristics of the depressive and non-depressive communities of LiveJournal.

Category	Community Name	#Members	#Posts	Description of the community
Depressive	alonendepressed	181	401	This community is for alone people suffering from depression.
	depression_uk	107	400	This is a community primarily for the discussion/support of depression sufferers in the UK.
	fightdepression	201	400	This community is meant to help those with depression.
	depressedteens	184	377	This is a community for depressed teenagers.
	imissmydad	163	441	This community supports people who have lost their fathers.
	parenting101	162	396	This community is for advice/personal experiences from many different types of parents.
Non-Depressive	computerhelp	225	401	This community provides free technical support for computer users.
	beauty101	219	410	This community helps to find answers of questions related to beauty.
	burning-man	186	400	This community supports LiveJournal Camp @ Burning Man
	dear-you	209	400	A place for unsent letters

4.3. Experiments and Results

To evaluate the performance of post classification, community classification, and depression degree classification, the dataset was split according to 10-fold cross-validation approach.

For post classification, the algorithm was trained on 3,626 posts and tested on 200 depressive posts and 200 non-depressive posts for each fold of the experiment. The confusion matrix of post classification is shown in Table 4. The results show that ratio of misclassification for depressive posts is lower in comparison to misclassified non-depressive posts.

For community classification, in each fold of the 10-fold cross-validation experiment, 10 communities were arbitrarily constructed with each community taking 200 posts each from the depressive and non-depressive posts. We used the leave one community out approach for both depressive and non-depressive classes. Thus, the algorithm was trained on 18 communities and tested on the remaining two communities for each fold of the experiment. For the computation of results, we considered a vote for each post as depressive and non-depressive and used the majority voting for assigning a final class for a community as depressive or non-depressive. The confusion matrix for community classification is shown in Table 5, which shows that there is not a single misclassified depressive community and only one non-depressive community is classified as a depressive community.

Table 4. The confusion matrix for post classification.

		Predicted	
		Non-depressive	Depressive
Actual	Non-depressive	1787	220
	Depressive	190	1829

Table 5. The confusion matrix for community classification.

		Predicted	
		Non-depressive	Depressive
Actual	Non-depressive	9	1
	Depressive	0	10

Random forest (RF) is a powerful classifier with the natural ability to build an accurate model for multi-class classification. It is a computationally efficient algorithm credited to work well for a variety of classification problems. We compare the proposed RF based method with Support Vector Machines (SVM) [43] classifier, another well-known technique for text classification which is based on finding the maximum margin between the classes. The precision, recall, f-measure and accuracy of the post and community classification is shown in Table 6. For both post and community classification, RF performs better in comparison to SVM. The proposed approach achieved about 90% and 95% accuracy in classifying the depressive posts and depressive communities, respectively. The performance of RF classifier is favourably higher than the SVM for both post and community classification, as seen in Table 6.

Table 6. Precision, Recall, F-Measure and Accuracy for post and community classification.

Classification Model	Classification	Precision	Recall	F-Measure	Accuracy
Random Forests	Post Classification	0.892	0.905	0.897	0.898
	Community Classification	0.900	1.000	0.947	0.950
SVM	Post Classification	0.818	0.783	0.799	0.820
	Community Classification	0.875	0.885	0.879	0.895

Table 7. The confusion matrix for depression degree classification.

		Predicted		
		Severe Depression	Moderate Depression	Mild Depression
Actual	Severe Depression	232	18	-
	Moderate Depression	11	224	23
	Mild Depression	-	17	283

For depression degree classification, only depressive posts were considered. First, the existing mood tags of the posts were mapped to either severe, moderate, or mild depression. Later, for the posts without mood tags, the mood tags of the post were automatically predicted as per the method described in Section 3.5.

We extracted 800 depressive posts with 250 posts each for severe and moderate depression and 300 posts for mild depression. Similar to the previous experiments, 10-fold cross-validation was followed to avoid any biasedness. The results of depression degree classification are shown in Table 7. The precision, recall, f-measure and accuracy of depression degree classification is shown in Table 8. The classification accuracy of the proposed algorithm is 92%. These results show the good performance of the proposed method for depression degree classification.

Table 8. Precision, Recall, F-Measure and Accuracy for depression degree classification.

Classification	Precision	Recall	F-Measure	Accuracy
Depression Degree Classification	0.927	0.922	0.924	0.923

4.4. Discussion

The post classification and community classification results from Tables 4, 5, and 6 demonstrate the promise of the proposed method which provided a high degree of precision, recall, and accuracy. These results show that the ratio of misclassification for depressive posts and depressive communities is lower in comparison to misclassified non-depressive posts and non-depressive communities, respectively. This is due to the reason that depressive communities mostly contain posts relevant to depression. However, non-depressive community users discuss various aspects of life and sometimes express their feelings of depression and thus potentially making the post a candidate for depressive class. The classification of such outliers is indeed a big challenge. Based on passive writing style and sentiments of these apparently non-depressive posts from non-depressive communities, the classifier labelled them as depressive posts that contradicted with the ground truth of the dataset.

Figure 4. Word cloud for moods of depressive posts.

Tables 7 and 8 illustrate the high accuracy results obtained for depression degree classification to predict the severity of depression for depressive posts. It is observed from the confusion matrix in Table 7 that the proposed method is able to discriminate severe and mild depression with very high accuracy without confusing these two levels of depression. However, some cases of moderate depression are observed to be misclassified as either severe or mild depression. This

can be attributed to the fact that relatively smaller difference exists between the valence of words in these classes since moderate depression has overlap in its descriptors with both severe and mild depression cases. It is a big challenge to accurately discriminate moderate depression cases from severe or mild depression cases and this can benefit from development of strongly discriminant descriptors in future work.

The relative frequency of words, used as mood tags in the depressive posts is shown in word cloud in Figure 4, where bigger words denote higher frequency. A word cloud visualization of both depressive and non-depressive communities is shown in Figure 5(a) and Figure 5(b), respectively. It is observed from these word clouds that the linguistic style of depressive content was considerably different from non-depressive content. The depressive posts contained more self-focused attention words in comparison to non-depressive posts. These word clouds depict that the depressive posts frequently contain words with depressive connotations which can be exploited by automatic prediction systems similar to the one proposed in this work. Such a system can facilitate the identification of users with depression symptoms at an early stage in order to avoid untoward incidents.

This work demonstrated the use of linguistic analysis and sentiment analysis along with machine learning to discriminate depressive content from non-depressive content. However, the proposed approach has applications in other areas of user-generated content analysis on social media platforms. This may include analysis of social media communities like sports, religion, technology, news, and other categories.

This work considered all depression categories as a single class to classify posts or communities. Future work will explore the classification of posts or communities in to depression category such as bipolar disorder, seasonal affective disorder, and postpartum depression before identifying the degree of depression.

(a). Word cloud for depressive communities.   **(b).** Word cloud for non-depressive communities.

Figure 5. Word cloud for depressive and non-depressive communities.

5. Conclusion

This paper presented a system which is able to accurately classify social media posts and communities in to depressive or non-depressive classes. For each depressive post, the proposed system can further determine the severity of depression of the user-generated content. The proposed system enabled the utilization of LIWC as a text analysis tool to convert LIWC data dimensions output into effective predictors. These discriminant predictors were employed by random forests classifier to accurately identify the depressive posts and communities from non-depressive ones. For each depressive post, the proposed method predicted the degree of depression (severe, moderate or mild) with high success on the basis of valence values for posts containing mood tags and using HMM for posts without mood tags. The experimental evaluation on dataset from LiveJournal community portal demonstrated the success of the proposed method achieving high classification accuracy of 90%, 95% and 92% for depressive posts, depressive communities and depression degree, respectively. The

presented results clearly illustrated the predictive capability of the proposed system to efficiently identify the depressive posts, depressive communities and depression degree from user-generated content in online social communities.

Notes

1. https://www.ncbi.nlm.nih.gov/books/NBK44246/
2. http://www.dhhs.tas.gov.au/mentalhealth/about_mental_illness/types_of_mental_illness
3. http://www.livejournal.com/

Funding

References

[1] Coppersmith G, Dredze M, Harman C, and Hollingshead K. From ADHD to SAD: Analyzing the language of mental health on Twitter through self-reported diagnoses. In: Proceedings of the 2nd workshop on computational linguistics and clinical psychology: From linguistic signal to clinical reality. Denver, Colorado: 2015, p. 1-10.

[2] Kumar M, Dredze M, Coppersmith G. and Choudhury MD. Detecting changes in suicide content manifested in social media following celebrity suicides. In: Proceedings of the 26th ACM conference on hypertext and social media, Northern Cyprus: 2015, p. 85-94.

[3] Wykes T, Haro JM, Belli SR, and others. Mental health research priorities for Europe. The Lancet Psychiatry 2015; 2: 1036-1042.

[4] Whiteford HA, Degenhardt L, Rehm J, Baxter AJ, and others. Global burden of disease attributable to mental and substance use disorders: findings from the global burden of disease study 2010. The Lancet Psychiatry 2013; 382: 1575-1586.

[5] Brundtland GH. Mental Health: A call for action by world health ministers. World Health Organization, Geneva, 2014.

[6] Kazmer MM, Lustria MLA, Cortese J, Burnett G, Kim JH, Ma J, and Frost J. Distributed knowledge in an online patient support community: authority and discovery. Journal of the association for information science and technology 2014; 65.

[7] Choudhury MD, Gamon M, Counts S, and Horvitz E. Predicting depression via social media. In: Proceeding of AAAI conference on weblogs and social media: 2013.

[8] Park S, Lee SW, Kwak J, Cha M, and Jeong B. Activities on Facebook reveal the depressive state of users. Journal of medical Internet research, 2013; 15.

[9] Rodriguez AJ, Holleran SE, and Mehl MR. Reading between the lines: The lay assessment of subclinical depression from written delf-descriptions. Journal of personality 2010; 78:575-598.

[10] Keshtkar F, and Inkpen D. A hierarchical approach to mood classification in blogs. Natural Language Engineering 2012; 18: 61-81.

[11] Nguyen T, Phung D, Adams B, Tran T, and Venkatesh S. Classification and pattern discovery of mood in weblogs. In: Proceeding of Pacific-Asia conference on knowledge discovery and data mining, Berlin: 2010, p. 283-290.

[12] Pennebaker JW, Boyd RL, Jordan K, and Blackburn K. The development and psychometric properties of LIWC2015. The University of Texas at Austin research work, 2015.

[13] Van Dam NT, and Earleywine M. Validation of the Center for Epidemiologic Studies Depression Scale—Revised (CESD-R): Pragmatic depression assessment in the general population. Psychiatry Research 2011; 186: 128 - 132.

[14] Bagby RM, Ryder AG, Schuller DR, and Marshall MB. The Hamilton depression rating scale: has the gold standard become a lead weight? American Journal of Psychiatry 2004; 161: 2163-2177.

[15] Syn SY, and Oh S. Why do social network site users share information on Facebook and Twitter? Journal of Information Science 2015; 41: 553-569.

[16] Panahi S, Watson J, Partridge H. Information encountering on social media and tacit knowledge sharing. Journal of Information Science 2015; 42: 539 – 550.

[17] Nemati S, Reza A, Nilchi N. Incorporating social media comments in affective video retrieval. Journal of Information Science 2016; 42: 524-538.

[18] Liu Y, Lin Z, Zheng X, Chen D. Incorporating social information to perform diverse replier recommendation in question and answer communities. Journal of Information Science 2016; 42: 449-464.

[19] Choudhury MD. Can social media help us reason about mental health? In: Proceedings of the 23rd international conference on world wide web, Korea: 2014, p. 1243-1244.

[20] Oh S, Syn SY. Motivations for sharing information and social support in social media: A comparative analysis of Facebook, Twitter, Delicious, YouTube, and Flickr. Journal of the Association for Information Science and Technology 2015; 66: 2330-1643.

[21] Hwang, JD., and Kristy H. Crazy Mad Nutters: The Language of Mental Health. In: Proceedings of the 3rd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, San Diego, California, 2016, p. 52–62.

[22] Moreno MA, L. Jelenchick A, Egan KG, Cox E, Young H, Gannon KE, and Becker T. Feeling bad on Facebook: Depression disclosures by college students on a social networking site. Depression and Anxiety 2011; 28: 447-455.

[23] Thelwall Mike. TensiStrength: Stress and relaxation magnitude detection for social media texts. Journal of Information Processing & Management 2017; 53:106–121.

[24] Park M, McDonald DW, and Cha M. Perception differences between the depressed and non-depressed users in Twitter. In: Proceedings of the 7th international conference on weblogs and social Media: 2013.

[25] Savolainen R. Requesting and providing information in blogs and internet discussion forums. Journal of Documentation 2011; 67:863 – 886.

[26] Nguyen, T, Svetha V, and Dinh P. Textual Cues for Online Depression in Community and Personal Settings. In: Proceedings of 12th International Conference on Advanced Data Mining and Applications, ADMA, QLD, Australia, 2016, p. 12-15.

[27] Nguyen T, Phung D, Dao B, Venkatesh S, and Berk M. Affective and content analysis of online depression communities. IEEE Transactions on Affective Computing 2014; 5: 217-226.

[28] Bradley MM and Lang PJ. Affective Norms for English Words (ANEW): Instruction manual and affective ratings. The center for research in psychophysiology, University of Florida., Gainesville, 2010.

[29] Blei DM, Ng AY, and Jordan MI. Latent dirichlet allocation. Journal of Machine Learning Research 2003; 3: 993-1022.

[30] Malmasi S, Zampieri M, and Dras M. Predicting post severity in mental health forums In: Proceedings of the 3rd workshop on computational linguistics and clinical psychology: from linguistic signal to clinical reality, California: 2016, p.133-137.

[31] Saha B, Nguyen T, Phung D, and Venkatesh S. A Framework for classifying online mental health related communities with an interest in depression. Journal of Biomedical and Health Informatics 2016; 20: 1008 - 1015.

[32] Mishne G. Experiments with mood classification in blog posts. In: Proceedings of ACM SIGIR 2005 workshop on stylistic analysis of text for information access, 2005.

[33] Keshtkar F, and Inkpen D. Using sentiment orientation features for mood classification in blogs. In: Proceedings of the IEEE international conference on natural language processing and knowledge engineering: 2009, p. 1-6.

[34] Mishne G, and Rijke MD. Capturing global mood levels using blog posts. In: Proceedings of AAAI-CAAW-06, the spring symposia on computational approaches to analyzing weblogs. Stanford: 2006.

[35] Conglomerates P. Linguistic inquiry and word count-www.liwc.net [computer software], 2007.

[36] Robnik-Šikonja M, and Kononenko I. An adaptation of Relief for attribute estimation in regression. In: Proceeding of machine learning: fourteenth international conference. San Francisco: 1997, p. 296-304.

[37] Sun Y, and Wu D. A RELIEF based feature extraction algorithm. In: Proceedings of the 2008 SIAM international conference on data mining: 2008, p. 188-195.

[38] Ho TK. The random subspace method for constructing decision forests. IEEE Transactions on Pattern Analysis and Machine Intelligence 1998; 20: 832 – 844.

[39] Robnik-Šikonja M. Improving random forests. In: Proceeding of Machine Learning: ECML 2004: 15th European conference on machine learning. , Italy: 2004, p. 359--370.

[40] Mohammed M, and Gader P. Handwritten word recognition using segmentation-free hidden markov modeling and segmentation-based dynamic programming techniques. IEEE Transactions on Pattern Analysis and Machine Intelligence 1996; 18: 548—554.

[41] Kasteren TL, Englebienne G, and Krose BJ. An activity monitoring system for elderly care using generative and discriminative Models. Personal Ubiquitous Computing 2010; 14: 489—498.

[42] Fawcett T. An introduction to ROC analysis. Pattern Recognition Letters 2006; 27: 861–874.

[43] Duda RO, Hart PE, and Stork DG. Pattern Classification. 2nd Edition, Wiley, 2000.