Edinburgh Research Explorer

Examining the Role of Mood Patterns in Predicting Self-reported Depressive Symptoms

Examining the Role of Mood Patterns in Predicting Self-reported Depressive symptoms

Lucia Lushi Chen
Lushi.Chen@ed.ac.uk
School of Informatics, The University of Edinburgh
Edinburgh, United Kingdom

Walid Magdy
wmagdy@inf.ed.ac.uk
School of Informatics, The University of Edinburgh
Edinburgh, United Kingdom

Heather Whalley
heather.whalley@ed.ac.uk
Centre for Clinical Brain Sciences, The University of Edinburgh
Edinburgh, United Kingdom

Maria Wolters
maria.wolters@ed.ac.uk
School of Informatics, The University of Edinburgh
Edinburgh, United Kingdom

ABSTRACT

Researchers have explored automatic screening models as a quick way to identify potential risks of developing depressive symptoms. Most existing models often use a person's mood as reflected on social media at a single point in time as one of the predictive variables. In this paper, we study changes in mood over a period of one year using a *mood profile*, which explicitly models the changes of mood, and transitions between moods reflected on social media text. We used a subset of the "MyPersonality" Facebook data set that comprises users who have consented to and completed an assessment of depressive symptoms. The subset consists of 93,378 Facebook posts from 781 users. We observed less evidence of mood fluctuation expressed in social media text from those with low symptom measures compared to others with high symptom scores. Next, we leveraged a daily mood representation in Hidden Markov Models to determine its associations with subsequent self-reported symptoms. We found that individuals who have specific mood patterns are highly likely to have reported high depressive symptoms. However, not all of the high symptoms individuals necessarily displayed this characteristic, which indicates presence of potential subgroups driving these findings. Finally, we leveraged multiple mood representations to characterize levels of depression symptoms with a logistic regression model. Our findings support the claim that derived mood from social media text can be used as a proxy of real-life mood to infer depressive symptoms in the current sample. Combining the mood representations with other proxy signals can potentially advance the current automatic screening technology for research.

1 INTRODUCTION

Depression is the leading cause of disability worldwide. Initial efforts to detect depression signals from social media posts have shown promising results [2, 16, 18, 31, 32, 37, 54]. Given the high internal validity [18, 42], these automatic screening tests are potentially beneficial to clinical judgement. The existing models for automatic detection of depressive symptoms learn proxy diagnostic signals from social media data, such as help-seeking behaviour for mental health or medication names [16, 18]. In reality, individuals with depression typically experience depressed mood, lost of pleasure nearly in all the activities, feeling of worthlessness or guilt, diminished ability to think [4]. Therefore, a lot of the proxy signals used in these models lack the theoretical underpinnings for depressive symptoms and it is reported that the social media posts from many patients in the clinical setting do not contain these signals [20]. Based on this research gap, we propose to monitor a type of signal that is well-established as a class of symptom in affective disorders — mood. Mood is an experience of feeling that can last for hours, days or even weeks [4]. In this current work, we attempt to enrich current automatic screening technology for depression by constructing a 'mood profile' for social media users.

The variance of quality and intensity of mood and emotional reactions are referred to as "affective style" [17], which underlies one's risks of developing psychological disorders [1, 45]. Assessing affective style in everyday life is difficult in an experimental context because it requires a costly extended period of data collection. In

contrast, social media data contains longitudinal information that reflect one's emotional reactions to stimuli. Therefore, it can provide researchers with an alternative lens to examine the affective style of an individual, based on the premise that approval is obtained from social media users, and data privacy is well-protected.

Existing automatic screening models often include mood as a feature variable in the modeling process. However, there are a few methodological gaps in these models. First, most of them do not distinguish between mood and emotions. Emotion is a brief reaction to a specific stimulus, whereas mood has longer temporal duration [30]. Researchers using social media data to study mood or emotions often see a single post as reflecting mood [9, 12, 52]. However, a single social media post is likely to reflect a participant's emotions at the time rather than ongoing mood [6, 44]. In this current work, we adopted the definition of mood from Association et al. [3]: "mood is the pervasive and sustained 'emotional climate', and emotions are 'fluctuating changes in emotional 'weather' '". We sought to determine whether temporal mood representation derived from social media text is associated with (subsequent) self-reported depressive symptoms, (and if so,) what are the best approaches to represent mood as a time dependent variable (for future work?).

Furthermore, a majority of models in this line of research often ignore the fact that affect is inherently time dependent. Only a few models have adopted temporal affective patterns [18, 42]. Most of these models also formulate the associations between affect and depressive symptoms based on the frequencies of affective words used in social media text [14, 47]. They have neglected the fact that transitioning from one affective state to another also reflects symptoms of affective disorders [10, 23, 44, 48]. In this work, we explored and tested multiple approaches to represent the temporal affective patterns and the transitions of affective states.

Nevertheless, social media user often posts sporadically. The sparsity of social media data posits a big challenge in the modeling process. Most of the existing studies imputed missing values with the mean or simply removed users with a lower word count [18, 56] . Removing outliers is beneficial to the modeling process, however, the core symptom of major depressive disorder (MDD) is disinterest in social contact and social withdrawal [3], therefore posting sparsely might be one of the key/core symptoms from people with MDD. Removing the outliers might therefore result in removing those with severe symptoms from the sample. Therefore, it is necessary to use some modeling techniques to include the outliers.

Towards addressing the above methodological gaps described above, we designed multiple mood representations inferred from social media text with the following characteristics: (i) Temporal features (ii) Transitions from one mood state to another (iii) Posting behavior. Here we see all the mood representations as a *Mood Profile* for social media users. We formulate the following questions to explore the roles of mood in predicting depressive symptoms:

(1) Are mood representations derived from social media text associate with the severity of self-reported depressive symptoms?
(2) Which representation in the mood profile is most predictive of the severity of self-reported depressive symptoms?

Our main contributions in this study are:

(1) Constructing a mood profile for social media users based on their status updates. The mood profile encompasses representations that encoded the variance of mood intensity and alternations of mood states and the behavior of not posting.
(2) Examining the associations between the social media mood profile and users' depressive symptoms level.
(3) Examining which representation in the mood profile is more predictive to depressive symptoms level.

In our work, we analysed a set of 93,378 posts from 781 Facebook users (consented details). For each user, a mood profile is constructed based on the social media text. We found that people with low symptom level tend to have less fluctuations in the mood pattern derived from their social media text. We also modelled the mood representation with a Hidden Markov Model and we found the hidden states estimated based on the mood representation is highly related to depressive symptoms. Nevertheless, combining several representations in the mood profile is more (than which?) predictive to depressive symptom levels (f-score: 0.62). Our results suggest the mood profile derived from social media text can potentially serve as a reference for an individual's depressive symptom level. In addition, this technique could potentially advance the current automatic screening technology. The data-driven, evidential nature of our approach provides us with better insight into the relationship between mood and depression.

2 BACKGROUND

2.1 Depression and Mood

Moods are slow-moving states of feeling, influenced by others, objects or situations [45, 57]. The pattern of mood reflects one's vulnerability to developing affective disorders [44, 45]. Depressed mood is a symptom of mood disorders, such as major depressive disorder (characterized by a persistent feeling of sadness) and dysthymia (persistent mild depression) [4].

It is also well established that mood fluctuation and irritability are associated with many somatic and sensory dysfunctions. Frequent alternating between moods (typically a few days) and irregular cycles of mood underlie the behavioural features of a wide variety of conditions [1]. We expect to find similar associations between mood derived from social media text and depressive symptoms since existing studies have identified associations between emotions presented on social media data and symptoms mental illness. For example, many studies have established that participants with depressive symptoms use more negative affective words (e.g. sad, cry, hate) in their social media text than those who do not [18, 37].

2.2 Detect depressive symptoms with Sentiment

Studies which examine emotions derived from social media data often adopt sentiment analysis. This is a computational process that categorizes affect or opinions expressed in a piece of text. The extracted affect is called sentiment [35]. Most of the existing works use averaged sentiment over a long period of time (e.g. one year) as a feature to predict depressive symptoms [8, 16, 37, 54, 54, 56].

However, how the sentiment changes over time is an important aspect to infer affective disorders. Only a few studies have included sentiment as a time dependent feature in the model [18]. For example, De Choudhury et al. [18] used the momentum of the feature vector in the screening detection. Eichstaedt et al. [19] include temporal posting patterns, but not the temporal affect pattern. Chen et al. [15] using temporal measures of fine grained emotions to predict user's depression state. Recently, Reece et al. [42] adopted a Hidden Markov Model (HMM) to analyse the change of language in social media posts. Results showed that the shift of words in status updates indicate depression and (expand) PTSD symptoms. The above mentioned studies adopted a sliding window technique to define dynamic sentiment [15, 18, 42]. However, none of them objectively explored the size of time window and the slide increment. In addition, existing studies only focus on examining a continuous sentiment value. In this work, we used sentiment as a foundation to construct the mood profile by aggregating the sentiment in a sliding window. We also examine the changes of affective states, therefore, we included categorical sentiment value (e.g. positive, negative) in constructing the mood profile.

2.3 Posting Behavior and depressive symptoms

Social media users are known to communicate selectively due to self-presentation biases [27, 55]. They are less likely to reveal events that project negatively on themselves [29]. Social media users may be reluctant to share negative feelings and symptoms of depressed mood due to stigma and fear of potential repercussions. Self-representation biases leads to fundamental differences between real-life mood and social media mood.

Intuitively, people with severe depressive symptom levels would be expected to post less than people with fewer symptoms, however, studies examining the relationship between social media data and depressive symptom level often report an opposite association. Several studies found that individuals with a history of depression (determined from past medical history) tended to post more often compared with people without depression [49]. However, posting frequencies did not differ among patients with other conditions, such as hypertension, diabetes, headaches, back pain, anaemia, and cancer. Although some participants with depressive symptoms post more often, this might happen only when a participant is not severely depressed and less restrained by self-presentation biases or stigma. In this study, we see the behaviour of not posting as a variable in itself and observe if this behaviour has any predictive capacity with regards depressive symptoms.

3 DATA

For this study, the myPersonality data set [5, 58] was used. It contains Facebook posts of 180,000 participants collected from 2010 to 2012, enriched with a variety of additional validated scales [5]. The collection of myPersonality data complied with the terms of Facebook service, and informed consent for research use was obtained from all participants. Permission for the use of this database was obtained in 2018. Other publications using this dataset include [22, 50].

3.1 depressive symptom Screening Test

From the participants in the myPersonality data, we focused on 1047 participants who completed the Center for Epidemiologic Studies Depression Scale (CES-D). The CES-D is a 20 item scale that measures the presence of depressive symptoms in the general population [40]. It is one of the screening tests most widely used by health service provider. Radloff [40] proposed three groups of depression severity: low (0-15), mild to moderate (16-22), and high (23-60). For testing the accuracy of using mood patterns to predict self-reported depressive symptoms, we follow the practice from previous social media studies [18, 37, 42, 54] and adopted 22 as a cutoff point to divide participants into high symptoms and low symptom groups. This allows us to compare our model's performances with previous studies. However, we were additioanlly interested in a more nuanced picture of the mood changes in different symptom levels. Therefore, in the analysis of mood fluctuation, we follow the original study from Radloff [40] and divide participants into three groups using two cutoff points: 16 and 22.

The symptoms measured in CES-D include mood, anhedonia, the feeling of being worried, restless, changes in sleeping pattern and physical symptoms (such as lost of appetite) and irrational thoughts. The scale has been found to have high internal consistency, test-retest reliability [34, 40, 43], and validity [34].

3.2 Summary Statistics

Among the 1047 participants who completed the CES-D scale, we further removed 110 participants who were less than 18 years old. The CES-D survey was open from 2010 to December 2012, but MyPersonality only collected participants' status updates from Jan 2009 - Dec 2011. Since the status updates in 2012 were not available, we further removed participants who completed the scale in 2012 and who posted at least one post in the past year. Eventually we yielded a final set of 781 participants who had posted 93,378 posts over the past year before they took the test.

The average number of posts per user over one year was 120, this distribution was skewed by a small number of frequent posters, as evidenced by a median value of 73 posts per user (range). Figure 1 shows participants' count of posts up to one year before they completed the CES-D scale.

The mean age of the participant is 26 (sd = 11.7), 333 (43%) participants are male and 448 (57%) are female. Table 1 shows further details of the participants, including the ethnicity, gender and marital status.

Table 1: Demographic Information of the 781 Participants

Ethnicity	No.	%	Marital Status	No.	%
Black	38	4.3	Single	574	73.8
Asian Chinese	26	3.3	Divorced	28	3.5
Middle Eastern	13	1.7	Married	27	3.4
Native American	13	1.6	Married with Children	38	4
Other Asians	84	10.8	Partner	78	10
Not Specified	96	12.2	Not specified	36	4.5
White-American	309	39.2			
White-British	71	8.9			
White-Other	131	17.1			

Figure 1: Distribution of post count from participants
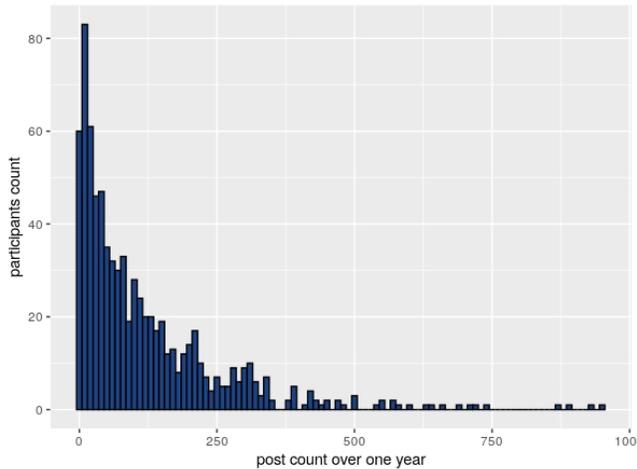
Note: Figure demonstrates the distribution of post count over one year before participants completed the CES-D survey scale. Size of the bin is 10.

Overall, our group of participants had a relatively high mean CES-D score (m = 26.3, sd = 8.9), see Figure 2. We found that the proportion of high symptom class to low symptom class in our sample is 1.6:1, which is high compared general population (ratio). Radloff [41] found only 21% of the general population scored at and above an arbitrary cutoff score of 16. However, we note the the current dataset is not an exceptional case. Other studies in this research area also obtained datasets with high symptom classes accounting for nearly half of the dataset. Reece et al. [42] used a dataset that contained 105 depressed participants and 99 non-depressed participants, other studies have a proportion of high symptom to low symptom class as 2:3 [18, 31, 54], 3:5 [33]. All of these studies recruited a sample biased towards potentially high symptom individuals compared with empirical studies which selected participants in a random trail. We speculate that there is a bias in those individuals self-selecting for this type of research.

Figure 2: Distribution of CES-D score
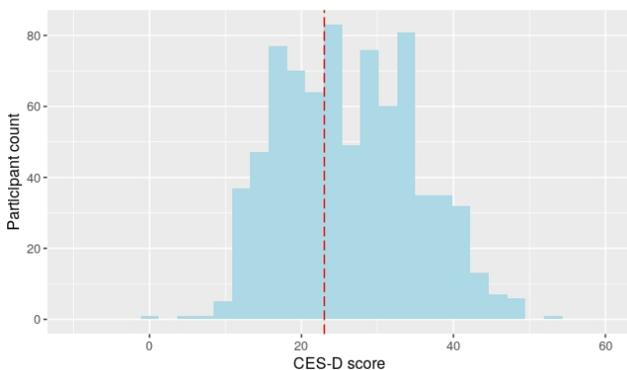
Note: Figure demonstrates the density distribution of the CES-D score, red line indicate the cutoff point 22

4 CONSTRUCTING MOOD PROFILE

A mood profile is constructed for each participant. Each mood profile encompasses sets of features which representing mood, the change of mood and the transition of mood states. Since mood is time dependent, we use a sliding window technique to construct the temporal features. A window starts from day 0 (the day when users completed the CES-D scale) and moves backwards for up to one year. Choosing the size of a time window presents a challenging question, how granular should a time window be? De Choudhury et al. [18] look at a user's tweet in a single day. Reece et al. [42] use both day and week as the time window because most of the participants did not generate enough daily content. In this paper, we define the size of the time window as measured by day $d \in D := \{3, 7, 14, 30\}$, see Table 2 for the notations. The size of the slide increment determines how much information the two adjacent windows share. The slide increment is also measured by day $s \in S := \{3, 7, 14\}$.

Another challenge is to decide how far back do social media posts indicate symptom level. Earlier studies use data up to one year before participants completed the self-reported symptom measurement [18], Reece et al. [42] found that symptoms can be predicted up to nine months before the onset of the illness. In the current work, each representation in the mood profile was constructed with posts written up to one year before participant completed depressive symptom survey.

Sentiment Scores. We used the sentiment scores retrieved from SentiStrength [53]. SentiStrength extracted sentiment from the text based on a function that describes how good the words and phrases of the text match a predefined set of sentiment lexicon.

Temporal Mood Representations. Since many social media users do not post every day, we encoded the behavior of not posting as "Silence" and we defined four mood states: positive, negative, neutral and silence. We adopted two approaches to define mood within a time window: most frequent mood state over a time window and average sentiment over a time window, see Table 2. For the average sentiment, silence days as missing values are imputed by the mean. We also constructed features that represent the change of mood [18], see mood momentum in Table 2.

Temporal Mood Transition Representations. We also encoded the probability of a user transferring from one mood state to another as a representation in the mood profile. We have in total 16 transition states with 16 combinations from the fours classes (positive, negative, neutral and silence), for example, from positive to negative, negative to silence. Note that if we set the slide increment as one day, we would have 365×16 mood transitions features. To prevent the large dimensionality, which might led to sparse representation, we defined d as 30 and s as 30, so that we have 12×16 feature columns for Mood Transition Representations.

5 ASSOCIATION BETWEEN MOOD PROFILE AND DEPRESSIVE SYMPTOMS

To examine whether the mood profile is associated with depressive symptoms, we observe whether its pattern is related to symptom level and test its predictive power on symptom level.

Table 2: Notations for Mood Profile

Variable	Notation	Description
Window Size	d	A period of time within $d \in D$ days, $d \in D := \{3, 7, 14, 30\}$
Slide Increment	s	A sliding window move forward by every $s \in S$ days, $s \in S := \{3, 7, 14\}$
Sentiment	v	Sentiment score of a single post
Day Sentiment	V	Arithmetic mean of sentiment in one day $V = \dfrac{v + \dots v_i}{i}$
$Mood_\mu$	M_μ	Arithmetic mean of day sentiment over a time window, $M_\mu = \dfrac{V + \dots V_d}{d}$
$Mood_\omega$	M_ω	Most frequent sentiment over a time window, categorical
Mood Momentum	ΔM	Difference between M_μ in two time windows
Mood States Transition	Tr	The probability of a user transfer from one mood-state to another, a mood state is defined by M_ω
Mood States Transition	ΔTr	Difference between Tr in two time windows

5.1 Mood Fluctuations

We modelled mood fluctuations using Gaussian Process (GP) regression. GP regression is a Bayesian approach that assumes a Gaussian process prior over functions [39]. In this analysis, we see the temporal mood representations as noisy representations of participants' mood. To reduce the noise, we use GP regression to estimate participants' latent mood based on their mood representation. For those participants with few data points, the GP regression is modeling the mean of the sample due to the imputation approach we adopted. Thus, for this experiment, we excluded participants posted less than 10 posts over year before they completed the depressive symptom scale. Eventually, this yielded 690 participants for the current analysis. We used mood representations with $d \in D := \{1, 3, 7, 14\}$ and $s \in S := \{1, 3, 7\}$ as input of the GP regression model. The smaller of d and s, the more noisy the data is. The GP regression is best fitted on mood vector with $d = 14$ and $s = 3$, see Figure 3. Each dot on the graph represents mood (averaged sentiment) in a time window $d = 14$, x axis shows the number of time windows d. Since the entirety of the data includes posts of one year (365 days), there are 122 time windows for each participant.

We constructed one model for each participant. Here we are not interested in making prediction with the GP regression model, we focus on the function parameter: lengthscale. The lengthscale describes how smooth a function it is, and small lengthscale means the function value changes quickly and vice versa [13]. By fitting a GP regression model on each user, we obtain a lengthscale of each user's latent mood, and we compare the lengthscale among participants with different symptom levels (low, moderate, high).

We used a nonparametric test (Mann-Whitney U test) to compare the lengthscale differences between groups. The lengthscale of the high symptom group (*Median* = 2.77) is identical to the moderate symptom (*Median* = 2.77) group ($U = 35424$, $p = 0.01$). However, the low symptom group (*Median* = 2.98) has a significantly larger

Figure 3: Example of GP Regression
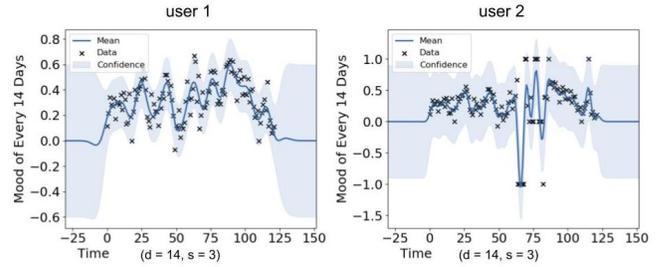
Note: here shows fitting GP regressions on four participants, each data point is mood of every 14 days estimated by the GP regression model. N = 781.

lengthscale than the high symptom group ($U = 17231$, $p = 0.01$). The moderate symptom group was also significantly different from the low symptom group ($U = 7244$, $p = 0.02$). Our result suggest that people with high or moderate depressive symptom level have more mood fluctuations than people with low symptom level.

5.2 Classifying Symptom Levels using Daily Mood Representation

Another approach to examine whether the mood profile is associated with depressive symptom is to examine whether participants having a particular mood state is influenced by depressive symptoms level. We assume the mood states are serially dependent and we used Hidden Markov Model (HMM) [7] to model two unobservable states based on a daily mood state representation. This representation comprises four mood states, including silence. Since the behavior of not posting (silence) is included in the modeling process, we did not remove any less active users in this analysis (N = 781).

Hidden Markov Model. We used a multinomial (discrete) emission Hidden Markov Model (HMM) to model the observed mood [26]. The major parameters used for the model are:

(1) Observed mood O_t (time series), daily mood transition representation ($d = 1$, $s = 1$).
(2) Transition matrix (A), gives the probability of a transition from one state to another.
(3) Transition state j.
(4) Observation emission matrix (B), which gives the probability of observing O_t when in state j.

An HMM model (denoted by λ) can be written as:

$$\lambda = (\pi, A, B) \tag{1}$$

The idea behind this approach is to use the observed mood to estimate the parameter set (π, A, B), A shows us the probability of transferring from one hidden state to another, and B tells us the probability of emitting a certain mood when a user is in a specific symptom state.

We used hmmlearn python library [24] to fit emission and transition matrices (using expectation-maximization) and hidden state sequence (using the Viterbi path algorithm), see Section A.1 for the initialized probabilities. We trained the model on the entire set of data and observed if the emission probabilities align with our

existing knowledge of affect and depressive symptoms. Here we are did not find the optimal model to forecast a new observation sequence, hence we did not test the training model on a test set. Instead, we decode a sequence of hidden states from the observations. We were interested to know whether the hidden states from the HMM model have anything to do with depressive symptoms.

The depressive symptoms measured in the CES-D scale include sleep disturbance, loss of appetite, feeling lonely, and a dozen psychological and physical signals. Most of these symptoms are associated with a depressed mood. An individual can experience more symptoms on some days and fewer symptoms on others. The HMM model decodes a binary hidden state for each day. Here we assume that one of the hidden states represents the user experiencing more depressive symptoms (high symptom state), and another represents fewer symptoms (low symptom state). We classify participants' symptom level according to the count of high symptom state. Here we use cutoff score 22 to divide participants into two groups for comparing the results with the existing models. However, there is a challenging questions, shall we count all the high symptom states over the entire one year? According to the CES-D scale, each participant was asked how many days they experienced any of the symptoms in the past week (e.g less than 1, 1-2, 3-4, 5 or more). Therefore, the hidden states sequence tracing back to one week or two weeks before participants completed the CES-D scale is more critical. Therefore, our classification criteria is to examine whether participants have at least x days experiencing high symptom in the last y days before they completed the CES-D scale, $x \in X := \{1, 2, 3, 4, 5, 6, 7\}$, $y \in Y := \{7, 14\}$.

5.2.1 Evaluation of Hidden States.

Emission Probabilities. We observed whether the hidden states' emission probabilities align with our existing knowledge in depressive symptom and affect. Table 3 shows two hidden states and their emission probabilities to each observation. Given an observed day, we can see both hidden states were most likely to emit silence day because social media users posted sparsely. However, the high symptom hidden state has lower probability to emit silence days compared with low symptom hidden state. The high symptoms state also has a higher probability to emit negative mood or neutral mood, but the low symptoms state has a higher probability to emit positive mood. Therefore, results from the HMM model aligns with our existing knowledge in depressive symptom and affect.

Table 3: Emission Probabilities

$N = 781$	Positive	Negative	Neutral	Silence
Low Symptom	8.51	5.20	4.65	81.6
High Symptom	3.15	12.8	7.00	76.9

Note: less symptoms: hidden state that represents less symptoms on a particular day, more symptoms: hidden state that represents more symptoms on a particular day, N: training sample size

Transition Probabilities of Observations. We are also interested to know whether people are more likely to transfer from certain mood states to another. We constructed a transition probability matrix for the observations (daily mood representation). Table 4 again shows

us that social media users in general are more likely to become silent after they posted any social media content, although high symptom group is less so. High symptom individuals have higher probabilities of changing in between any mood states other than silence. This result aligns with the findings from the GP regression that low symptom individuals shows less fluctuations in their mood representation.

In general, people were more likely to have a positive mood if they had a positive mood in the previous time window. The probabilities of $+ \rightarrow +$, $- \rightarrow -$ were similar among the two groups, but high symptom participants are slightly more likely to transfer from negative to negative. When low symptom participants have a neutral mood, they have similar chances of having a neutral or negative mood in the next time window, whereas, high symptom participants are also more likely to have a negative mood in the next time window. Our result shows that while people, in general, are more vocal when they have a negative mood, but high symptom participants are more likely to vocal about the negative content for a more extended period.

### Table 4: Transition Probabilities of Observations

	High symptom				Low symptom			
	+	-	0	S	+	-	0	S
+	21.1	15.7	13.4	49.6	19.5	13.3	12.3	54.8
-	22.3	16.2	14.1	47.3	20.5	13.3	12.9	53.3
0	19.3	14.5	12.8	53.3	17.6	11.6	11.8	58.9
S	5.82	37.5	4.21	86.2	5.92	37.1	4.33	85.9

Note: +: positive, −: negative, 0 neutral, S: silent

Using Hidden States to Classify Symptom Level. Figure 4 shows the precision and recall of the high symptom class by counting the hidden states from the HMM model. The baseline model is formulated using a stratified dummy classifier that predicts based on the most frequent training labels. Precision increases as the criterion of x increase. Table 5 shows some of the best classification results. Assigning participants with six high symptom states within 14 days to the high symptom class results in very low recall (10.8%) but high precision (71.2%). Assigning participants with one high symptom state within 14 days results in a more balanced recall (60.3%) and precision (58.1%) to high symptom class. Result from this classifier does not surpass the baseline in f1 score but when using a higher x as criteria, the precision rate is much higher than the baseline. Our result supports the claim that daily mood representations inferred from social media text is highly associated with depressive symptoms. When a social media user shows specific mood patterns, it is highly likely that the person developed high level of depressive symptoms. However, only using this approach to identify high symptom individuals would result in a lot of false negative cases.

6 REPRESENTATION PREDICTABILITY OF DEPRESSIVE SYMPTOMS

The previous analysis suggests that the mood profile is highly associated with depressive symptoms. Now we examine which representation in the mood profile is most predictive of depressive

Figure 4: Precision and Recall of High Symptom Class (HMM) with Various Assignment Criterion
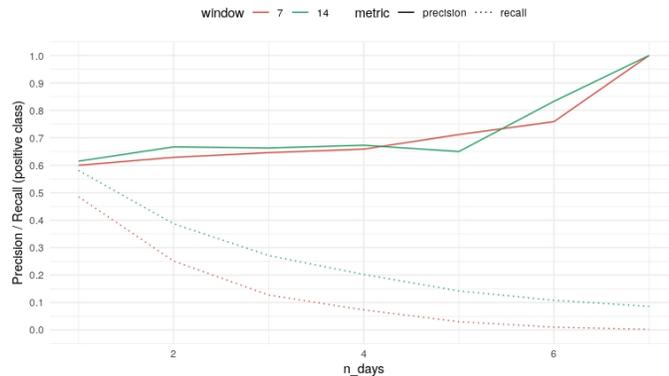
Note: window: size of the time window, x days before participants completed the CESD scale. ndays: count of high symptom state within the time window

Table 5: Predicting depressive symptom with hidden states

Criteria	P	R	f1
baseline	61.2	1.0	76.0
$x = 1, y = 7$	61.5	48.5	45.2
$x = 1, y = 14$	60.3	58.1	59.2
$x = 6, y = 14$	71.2	10.8	18.9

Note: high: high symptom class, low: low symptom class, R: recall of high symptom class, P: precision of high symptom class, f1: average macro-f1 score of both classes, criteria: criteria for classifying high symptom class

symptoms. We combine the representations with sets of proxy signals in a classification task.

6.1 Feature Extractionn

We extracted multiple features for the posts of each user to train multiple models for high-symptoms prediction. Our extracted featured included: 1) n-gram word representation, where $n \in N := \{1, 2, 3\}$); 2) topic modelling from Latent Dirichlet allocation (LDA) and 3) all the entries from Linguistic Inquiry and Word Count (LIWC) [38]. N-gram were ordered by term frequency across the corpus, we grid searched the number of most frequent n-gram and number of topics for LDA (see Section A.1). We found the most frequent 1500 n-grams and 30 LDA topics gave us the optimal results. These feature variables were commonly used in an automatic screening task [16, 18, 37, 42]. We compare the precision and recall between models with different representations from the mood profile.

Our dataset has an exceptionally high proportion of high symptom individuals as discussed earlier. Given that we have only 303 low symptom participants among 781 participants, we randomly selected 303 participants in the high symptom sample to have a dataset with a balanced class proportion that is closer to the existing literature (1:1), $N = 606$. We split the data into train (80%, N = 486), and test set (20%, N = 120) in stratified fashion. Stratified five-fold cross-validation was used to optimize the parameters in the model training. A grid search of parameters was carried out for

the logistic regression [51] classification algorithms, see Section A.1 for the grid search parameters.

6.2 Model Evaluation

A baseline model is formulated using a stratified dummy classifier that predicts based on the distribution of training labels. Out of several candidate algorithms (e.g. decision trees, support vector machine), logistic regression demonstrated best performance. Hundreds of classification models were trained and evaluated for this task. The models with different representations from the mood profile can be evaluated by precision and recall. We grid searched d and s that maximises the metrics. Figure 5 shows the precision and recall of the high symptom class from models with various configurations and feature sets. Models with configuration 4 (time window 30 days and increment slide 3 days) yield the best scores. Table 6 shows the precision, recall and f1 score of the high symptom class from configuration 4. The model with mood, mood momentum and mood transition representations yields the highest scores, and the model with averaged mood over a time window gives second highest scores, 0.59 precision, 0.65 recall, and an F-score of 0.62.

Figure 5: Precision and Recall of logistic regression
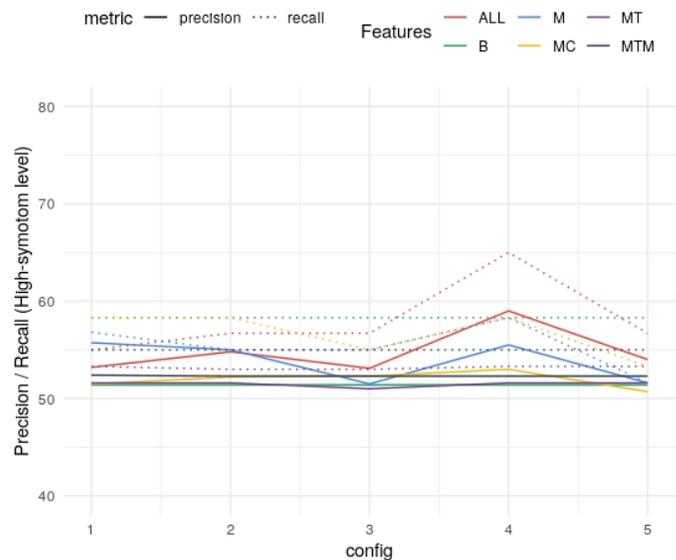
Note: config 1: $d = 7$, $s = 3$, config 2: $d = 14$, $s = 3$, config 3: $d = 14$, $s = 7$, config 4: $d = 30$, $s = 3$, config 5: $d = 30$, $s = 7$, B: basic features (n-gram, topic modeling, LIWC), M: B + $Mood_\mu$, MC: B + mood momentum, MT: B + mood transition, MTM: B + mood transition momentum, All: all features excluded MTM

7 DISCUSSION

7.1 The Role of Mood in Predicting Depressive symptoms

Mood is a time dependent variable, using time series approaches to model mood inferred from social media text provides us with better insight about mood and depressive symptoms. Participants demonstrated significantly fewer mood fluctuations when they reported a

Table 6: Prediction result of depressive symptom (d=14, s=3

Features	P	R	F1
RB	47.6	50.0	48.8
B	51.4	58.3	54.7
B + M_μ	55.5	58.3	56.9
B + Δm	53.0	58.3	55.6
B + Tr	51.6	53.3	52.4
B + ΔTr	52.3	55	53.6
B + M_μ + Δm + B + Tr	**59.0**	**65**	**61.9**

Note: R, P, F1 are recall, precision, and f1 score of high symptom classes respectively. B: basic features (tfidf bag-of-words, topic modeling, sentiment, LIWC). RB: random baseline, model parameters: penalty: l2, Inverse of regularization strength: 0.1)

low symptom score. This finding aligns with the well-established connection between emotionality and depression in the psychology literature. We also found the hidden states from the HMM model are highly relevant to self-reported depressive symptoms, see Table 5. Our model suggests that an individual having one high symptom state in 14 days is highly likely to have high symptom level. It is worth to note that the criteria we used in here is different from the criteria in the CES-D scale, which defines that an individuals experienced symptoms 1-2 days in the past 7 days might develop high symptom levels. However, using this approach alone to classify symptom level yielded to false negative classifications. This result suggests that individuals who show specific mood pattern in social media text are highly likely having high depressive symptoms, however, most of the individuals with high symptom do not display this mood pattern. We speculate the presence of potential subgroups driving these findings. Therefore, future studies could investigate the cause of this biases.

Existing studies that a use sliding window technique to create dynamic sentiment features haven't yet explored which representations and configurations tend to yield a better result in classifying symptoms. We explored various configurations of the sliding window and found that combining several representations in the mood profile together can dramatically enhance the model performance. Our best model (f-score: 0.62) encompasses the mood profile and a set of basic features commonly used in existing works. Other studies using multiple sets of proxy signals to predict depressive symptoms achieved a precision score ranging from 0.48 [16] to 0.87 [25, 42]. Schwartz et al. [46] using the same data set and achieved an accuracy of 0.386 (correlation) on continuous scores. The mood profile can potentially enhance the current screening technology by combining it with more advanced engineered features.

The transition probabilities of mood showed that participants, in general, were more vocal on social media when they were in a negative mood. The HMM emission probabilities also add to the claim that silent behaviour has implications for one's mental health. Participants with more symptoms were more vocal on social media, seemed to be contradicted with the fact that people with depression suffer from a loss of pleasure in usual activities. We speculate that some depressed individuals react to negative mood by posting, and some by silence. For those who react to negative mood by posting, they might be talking about their symptoms or reach out to others.

This sometimes were interpreted as social media platform making people depressed [36]. This finding together with the findings that some individuals with high symptom scores are associated with a specific mood pattern suggest heterogeneity and stratification.

Social media data provides noisy signals for users' mood. Despite the fact that social media users tend to be subject to positive self-presentation bias and they often post sporadically, our results show that a temporal mood profile derived from social media text is highly associated with users' subsequent self reported depressive symptom level. In particular, mood momentum and mood transition states may potentially enhance the current screening technology for research, especially applying advance time series technique on these representations. Most importantly, this mood profile can potentially provide more information to clinicians than a classification system with binary output.

7.2 Technological and Ethical Implications

A recent study suggests that incorporating clinical judgment via an appraisal of social media self-reports of mental illnesses leads to the best performance [20]. However, this finding only relevant for those people who disclose or discuss their condition on Social Media. Similar to the existing studies, the present finding of the derived mood pattern has implications on symptom level but does not provide an accurate interpretation for participants' mental health condition. An accurate interpretation of one's mental health condition must be involved looking at multiple perspectives of a person's life. The daily life information contained in social media data is just a tip of the iceberg of one's life experience, anyone who is using a similar approach on symptom inference should interpret the symptoms while combining it with real-life history. Therefore, the desired outcome of using this technique should be to identify individuals with mood abnormality that might expose one to a certain level of vulnerability to develop affective symptoms. Participants need to be informed of their scores and a diagnosis result should be a collaborative work from both the patient and providers.

Our approaches provide a useful source of information for assessing participants' derived mood pattern over time. However, data privacy and ethical research practices should be the top priority concern, given recent scandals about misusing social media data [11, 21, 28]. Using such a technique must be based on the premise that consent is obtained from participants and participants are comfortable adopted such techniques to assess their mental health status. Future research can prioritize assessing public opinion on using these techniques and explore how to build up confidence in data privacy and data protection.

7.3 Limitations

Our sample contains participants who allowed researchers access to their Facebook posts and to complete a symptom screening scale. Therefore, this sample may be strongly biased towards those who were comfortable to disclose and reach out on social media. It is still unclear about what the biases are in a sample with these tendencies compared with a random patient sample. Of particular interests is the relatively high depressive symptom score from most of the participants in this sample, and this bias is prevalent in studies

of this line of research [25]. We speculated that people who have depression are more curious about taking part in mental health related studies.

In this work, the symptom screening test was conducted once only. There were also no tests controlling for the presence of other disorders, such as bipolar, which greatly affect behaviour and mood variability. Those at the high end of the scale could have other types of affective disorders but showing depressive symptoms at the time when they carried out the self-reported measurement. Therefore, the measurement of self-reported symptoms is not an accurate reflection of whether the person has depression.

In addition to that, the sentiment scores employed in this study were retrieved with SentiStrength, which is a word counting approach to identify positive and negative affect. Although numerous studies have validated the word counting approach, the ideal method to retrieve less noisy sentiment is to construct the sentiment classification model with the examined dataset. Future studies can train their model for sentiment annotation to retrieve more accurate sentiment.

8 CONCLUSION

Mood is an important signal for the development of a depression episode. This report provides an outline of utilizing the sliding window technique to construct temporal representations of mood based on sentiment expressed in social media text. The behavior of not posting was also encoded in some of the representations. However, mood inferred from social media text is different from mood in real life. In order to examine whether the mood profile inferred from social media text is associated with depressive symptoms, we use the mood profile to classify depressive symptom level with time-series modelling and logistic regression algorithm. Our result suggests that the mood profile inferred from social media data is highly predictive of depressive symptoms, especially when the behavior of not posting is included. We also discover a pattern whereby people are more vocal in social media when they are unhappy. Despite many social media users being subject to positive self-presentation biases, social media provides a place for people to channel their emotions. Future studies can focus on studying this behavior on an actual patient group and a random control sample. The techniques proposed here offer a novel contribution to the current automatic screening technology as they are not focused on providing a binary classification result, but a longitudinal reference for the development of depressive symptoms.

A EXPERIMENTAL DETAILS

The following supplementary material details what is required to reproduce our results as closely as possible.

A.1 MODEL TRAINING

Grid searches of the following pairings of parameter spaces and Scikit-Learn implementations of algorithms were carried out:

- Feature Extraction
  - number of n-gram: 1000, 1500, 3000, 4000, 5000, 6000
  - number of topics: 10, 20, 30
- HMM:
  - Initial transition probability: [0.5, 0.5], [0.5, 0.5]
  - Initial transition probability: [0.2, 0.3, 0,2, 0.3], [0.2, 0.2, 0.3, 0.3]
  - Number of iteration: 10
- Support Vector Machine
  - Inverse of regularization strength: 0.5, 0.7, 1.0, 1.5, 2.0, 2.5
  - Kernel: linear, poly, rbf, sigmoid
  - Kernel coefficient: 0.01, 0.001, 0.0005
- Extra Trees
  - Number of Estimators: 100, 300, 500, 1000
  - Maximum Tree Depth: 20, 50, 100, 200
  - Maximum number of features: sqrt, log2
- Logistic Regression:
  - Penalty: l1, l2
  - Inverse of regularization strength: 0.1, 0.3, 0.5, 0.7, 0.9, 1.0, 1.5, 2.0

REFERENCES

[1] Hagop S Akiskal. 1996. The temperamental foundations of affective disorders. *Interpersonal factors in the origin and course of affective disorders* (1996), 3–30.

[2] Hayda Almeida, Antoine Briand, and Marie-Jean Meurs. 2017. Detecting Early Risk of Depression from Social Media User-generated Content.. In *CLEF (Working Notes)*.

[3] American Psychiatric Association et al. 2000. Diagnostic and statistical manual-text revision (DSM-IV-TR). *Washington, DC: American Psychiatric Association* (2000), 28–819.

[4] American Psychiatric Association et al. 2013. *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub.

[5] Yoram Bachrach, Michal Kosinski, Thore Graepel, Pushmeet Kohli, and David Stillwell. 2012. Personality and patterns of Facebook usage. In *Proceedings of the 4th annual ACM web science conference*. ACM, 24–32.

[6] C Daniel Batson, Laura L Shaw, and Kathryn C Oleson. 1992. Differentiating affect, mood, and emotion: toward functionally based conceptual distinctions. (1992).

[7] Matthew J Beal, Zoubin Ghahramani, and Carl E Rasmussen. 2002. The infinite hidden Markov model. In *Advances in neural information processing systems*. 577–584.

[8] Adrian Benton, Margaret Mitchell, and Dirk Hovy. 2017. Multi-task learning for mental health using social media text. *arXiv preprint arXiv:1712.03538* (2017).

[9] Johan Bollen, Huina Mao, and Alberto Pepe. 2011. Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. In *Fifth International AAAI Conference on Weblogs and Social Media*.

[10] Lauren M Bylsma, April Taylor-Clift, and Jonathan Rottenberg. 2011. Emotional reactivity to daily events in major and minor depression. *Journal of abnormal psychology* 120, 1 (2011), 155.

[11] Carole Cadwalladr and Emma Graham-Harrison. 2018. Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. *The guardian* 17 (2018), 22.

[12] Fabio Celli, Arindam Ghosh, Firoj Alam, and Giuseppe Riccardi. 2016. In the mood for sharing contents: Emotions, personality and interaction styles in the diffusion of news. *Information Processing & Management* 52, 1 (2016), 93–98.

[13] Krzysztof Chalupka, Christopher KI Williams, and Iain Murray. 2013. A framework for evaluating approximation methods for Gaussian process regression. *Journal of Machine Learning Research* 14, Feb (2013), 333–350.

[14] Lushi Chen, Christopher Hon Kwong Cheng, and Tao Gong. 2020. Inspecting Vulnerability to Depression from Social Media Affect. *Frontiers in Psychiatry* 11 (2020), 54.

[15] Xuetong Chen, Martin D Sykora, Thomas W Jackson, and Suzanne Elayan. 2018. What about mood swings: identifying depression on twitter with temporal measures of emotions. In *Companion Proceedings of the The Web Conference 2018*. 1653–1660.

[16] Glen Coppersmith, Mark Dredze, and Craig Harman. 2014. byt. In *Proceedings of the workshop on computational linguistics and clinical psychology: From linguistic signal to clinical reality*. 51–60.

[17] Richard J Davidson. 1998. Affective style and affective disorders: Perspectives from affective neuroscience. *Cognition & Emotion* 12, 3 (1998), 307–330.

[18] Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. 2013. Predicting depression via social media. In *Seventh international AAAI conference on weblogs and social media*.

[19] Johannes C Eichstaedt, Robert J Smith, Raina M Merchant, Lyle H Ungar, Patrick Crutchley, Daniel Preoţiuc-Pietro, David A Asch, and H Andrew Schwartz. 2018. Facebook language predicts depression in medical records. *Proceedings of the*

National Academy of Sciences 115, 44 (2018), 11203–11208.

[20] Sindhu Kiranmai Ernala, Michael L Birnbaum, Kristin A Candan, Asra F Rizvi, William A Sterling, John M Kane, and Munmun De Choudhury. 2019. Methodological gaps in predicting mental health states from social media: Triangulating diagnostic signals. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 134.

[21] Susan T Fiske and Robert M Hauser. 2014. Protecting human research participants in the age of big data.

[22] Jan-Philipp Freudenstein, Christoph Strauch, Patrick Mussel, and Matthias Ziegler. 2019. Four personality types may be neither robust nor exhaustive. *Nature human behaviour* 3, 10 (2019), 1045–1046.

[23] Nico H Frijda. 1993. Moods, emotion episodes, and emotions. *Handbook of emotions* 12, 2 (1993), 155.

[24] Zhong-Ke Gao, Michael Small, and Juergen Kurths. 2017. Complex network analysis of time series. *EPL (Europhysics Letters)* 116, 5 (2017), 50001.

[25] Sharath Chandra Guntuku, David B Yaden, Margaret L Kern, Lyle H Ungar, and Johannes C Eichstaedt. 2017. Detecting depression and mental illness on social media: an integrative review. *Current Opinion in Behavioral Sciences* 18 (2017), 43–49.

[26] Mathias Johansson and Tomas Olofsson. 2007. Bayesian model selection for Markov, hidden Markov, and multinomial models. *IEEE signal processing letters* 14, 2 (2007), 129–132.

[27] Junghyun Kim and Jong-Eun Roselyn Lee. 2011. The Facebook paths to happiness: Effects of the number of Facebook friends and self-presentation on subjective well-being. *CyberPsychology, behavior, and social networking* 14, 6 (2011), 359–364.

[28] D Lumb. 2016. Scientists release personal data for 70,000 OkCupid profiles. *Available at engt. co/2b4NnQ0. Accessed August 15* (2016), 2016.

[29] Soraya Mehdizadeh. 2010. Self-presentation 2.0: Narcissism and self-esteem on Facebook. *Cyberpsychology, behavior, and social networking* 13, 4 (2010), 357–364.

[30] William N Morris. 2012. *Mood: The frame of mind*. Springer Science & Business Media.

[31] Moin Nadeem. 2016. Identifying depression on Twitter. *arXiv preprint arXiv:1607.07384* (2016).

[32] Thin Nguyen, Dinh Phung, Bo Dao, Svetha Venkatesh, and Michael Berk. 2014. Affective and content analysis of online depression communities. *IEEE Transactions on Affective Computing* 5, 3 (2014), 217–226.

[33] Ahmed Husseini Orabi, Prasadith Buddhitha, Mahmoud Husseini Orabi, and Diana Inkpen. 2018. Deep learning for depression detection of twitter users. In *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*. 88–97.

[34] John G. Orme, Janet Reis, and Elicia J. Herz. 1986. Factorial and discriminant validity of the center for epidemiological studies depression (CES-D) scale. *Journal of clinical psychology* 42, 1 (1986), 28–33.

[35] Bo Pang, Lillian Lee, et al. 2008. Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval* 2, 1–2 (2008), 1–135.

[36] Igor Pantic. 2014. Online social networking and mental health. *Cyberpsychology, Behavior, and Social Networking* 17, 10 (2014), 652–657.

[37] Minsu Park, Chiyoung Cha, and Meeyoung Cha. 2012. Depressive moods of users portrayed in Twitter. In *Proceedings of the ACM SIGKDD Workshop on healthcare informatics (HI-KDD)*, Vol. 2012. 1–8.

[38] James W Pennebaker, Martha E Francis, and Roger J Booth. 2001. Linguistic inquiry and word count: LIWC 2001. *Mahway: Lawrence Erlbaum Associates* 71, 2001 (2001), 2001.

[39] Joaquin Quiñonero-Candela and Carl Edward Rasmussen. 2005. A unifying view of sparse approximate Gaussian process regression. *Journal of Machine Learning Research* 6, Dec (2005), 1939–1959.

[40] Lenore Sawyer Radloff. 1977. The CES-D scale: A self-report depression scale for research in the general population. *Applied psychological measurement* 1, 3 (1977), 385–401.

[41] Lenore Sawyer Radloff. 1977. The CES-D scale: A self-report depression scale for research in the general population. *Applied psychological measurement* 1, 3 (1977), 385–401.

[42] Andrew G Reece, Andrew J Reagan, Katharina LM Lix, Peter Sheridan Dodds, Christopher M Danforth, and Ellen J Langer. 2017. Forecasting the onset and course of mental illness with Twitter data. *Scientific reports* 7, 1 (2017), 13006.

[43] Robert E. Roberts. 1980. Reliability of the CES-D scale in different ethnic contexts. *Psychiatry research* 2, 2 (1980), 125–134.

[44] Jonathan Rottenberg. 2005. Mood and emotion in major depression. *Current Directions in Psychological Science* 14, 3 (2005), 167–170.

[45] Jonathan Rottenberg and James J Gross. 2003. When emotion goes wrong: Realizing the promise of affective science. *Clinical Psychology: Science and Practice* 10, 2 (2003), 227–232.

[46] H Andrew Schwartz, Johannes Eichstaedt, Margaret Kern, Gregory Park, Maarten Sap, David Stillwell, Michal Kosinski, and Lyle Ungar. 2014. Towards assessing changes in degree of depression through facebook. In *Proceedings of the workshop on computational linguistics and clinical psychology: from linguistic signal to clinical reality*. 118–125.

[47] H Andrew Schwartz, Johannes C Eichstaedt, Margaret L Kern, Lukasz Dziurzynski, Stephanie M Ramones, Megha Agrawal, Achal Shah, Michal Kosinski, David Stillwell, Martin EP Seligman, et al. 2013. Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS one* 8, 9 (2013), 234.

[48] Gal Sheppes, Gaurav Suri, and James J Gross. 2015. Emotion regulation and psychopathology. *Annual review of clinical psychology* 11 (2015), 379–405.

[49] Robert J Smith, Patrick Crutchley, H Andrew Schwartz, Lyle Ungar, Frances Shofer, Kevin A Padrez, and Raina M Merchant. 2017. Variations in facebook posting patterns across validated patient health conditions: a prospective cohort study. *Journal of medical Internet research* 19, 1 (2017), e7.

[50] Xiangguo Sun, Bo Liu, Qing Meng, Jiuxin Cao, Junzhou Luo, and Hongzhi Yin. 2019. Group-level personality detection based on text generated networks. *World Wide Web* (2019), 1–20.

[51] Johan AK Suykens and Joos Vandewalle. 1999. Least squares support vector machine classifiers. *Neural processing letters* 9, 3 (1999), 293–300.

[52] Mike Thelwall, Kevan Buckley, and Georgios Paltoglou. 2011. Sentiment in Twitter events. *Journal of the American Society for Information Science and Technology* 62, 2 (2011), 406–418.

[53] Mike Thelwall, Kevan Buckley, Georgios Paltoglou, Di Cai, and Arvid Kappas. 2010. Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology* 61, 12 (2010), 2544–2558.

[54] Sho Tsugawa, Yusuke Kikuchi, Fumio Kishino, Kosuke Nakajima, Yuichi Itoh, and Hiroyuki Ohsaki. 2015. Recognizing depression from twitter activity. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. ACM, 3187–3196.

[55] Erin A Vogel, Jason P Rose, Lindsay R Roberts, and Katheryn Eckles. 2014. Social comparison, social media, and self-esteem. *Psychology of Popular Media Culture* 3, 4 (2014), 206.

[56] Xinyu Wang, Chunhong Zhang, Yang Ji, Li Sun, Leijia Wu, and Zhana Bao. 2013. A depression detection model based on sentiment analysis in micro-blog social network. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, 201–213.

[57] David Watson. 2000. *Mood and temperament*. Guilford Press.

[58] Wu Youyou, Michal Kosinski, and David Stillwell. 2015. Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences* 112, 4 (2015), 1036–1040.