

MentalSpot: Effective Early Screening for Depression Based on Social Contagion

Jahandad Pirayesh
California State University,
Sacramento
jpirayesh@csus.edu

Haiquan Chen
California State University,
Sacramento
haiquan.chen@csus.edu

Xiao Qin
Auburn University
xqin@auburn.edu

Wei-Shinn Ku
Auburn University
weishinn@auburn.edu

Da Yan
University of Alabama at Birmingham
yanda@uab.edu

ABSTRACT

While depression is rated as the most important leading factor to global disability, early detection of depression is a non-trivial task. Existing depression detection mechanisms harvesting social media data suffer from two major limitations. First, existing solutions rely heavily on the amount, quality, and variety of content types (textual, visual, etc.) posted by users to make accurate inferences, therefore suffering from the cold-start problem when coping with users with limited training data (e.g., most existing works exclude users with fewer than 25 tweets). Second, existing approaches ignore the social impact or indication from users' social circles that can be leveraged to enhance the inference results. In this paper, we present MentalSpot, a social-contagion based depression early-screening framework using meta-learning. Specifically, we first construct a social-contagion driven data repository PsycheNet, filling the void of social-circle based depression datasets. We design a triplet network to extract users' embeddings based on the similarities of the linguistic features extracted from written texts. Afterwards, for each target user, we employ dynamic mean shift pruning to select her top- k homogeneous friends in the metric space, the texts written by whom will then be leveraged to train a friend based depression detection model. Extensive experiments show that MentalSpot outperforms the state of the art in terms of all effectiveness metrics, especially for users with very few tweets. Specifically, by using only five tweets per user, MentalSpot successfully yields an F1 score that would otherwise be achieved by the state-of-the-art methods requiring at least twenty tweets. Our approach represents a step forward to address the cold-start problem that deep learning techniques struggle with for their applications in psychiatric diagnosis. The principal beneficiaries of this study are healthcare professionals in medical institutions to determine timely and targeted interventions in a clinical setting. This study also supports non-profit groups in reaching out to people with mental health issues, helping in a global health task that cannot be fully covered by clinicians.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CIKM '21, November 1–5, 2021, Virtual Event, QLD, Australia

© 2021 Association for Computing Machinery.

ACM ISBN 978-1-4503-8446-9/21/11...\$15.00

<https://doi.org/10.1145/3459637.3482366>

CCS CONCEPTS

• **Computing methodologies** → **Machine learning**.

KEYWORDS

Meta-Learning; Neural Embedding; Depression; Early Detection

ACM Reference Format:

Jahandad Pirayesh, Haiquan Chen, Xiao Qin, Wei-Shinn Ku, and Da Yan. 2021. MentalSpot: Effective Early Screening for Depression Based on Social Contagion. In *Proceedings of the 30th ACM International Conference on Information and Knowledge Management (CIKM '21), November 1–5, 2021, Virtual Event, QLD, Australia*. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3459637.3482366>

1 INTRODUCTION

Clinical depression, also known as Major Depressive Disorder (MDD), is a significant medical condition marked by persistent feelings of emptiness, sadness, and loneliness. Depression affects not only individuals but also the society and economy. According to the World Health Organization, depression is one of the leading causes of disability worldwide affecting more than 300 million people globally. Also, it is estimated that the lost productivity associated with depression and anxiety disorders costs the global economy more than 1 trillion dollars each year [17]. Even though depression can be treated, it often goes untreated because of existing stigma, shame, and financial barriers. Undiagnosed and untreated depression is often associated with an increased risk of substance abuse, self-harm, and suicide, which is the second leading cause of death among individuals between the ages of 10 and 34 [8].

Clinical Limitations of Existing Approaches. Traditional classification models such as Support Vector Machine, k -Nearest Neighbor, Logistic Regression, and Random Forest were explored for depression detection [4, 6, 15, 19, 20]. Over the last decade, deep learning techniques have gained popularity in depression detection [11, 14, 33, 35] because of their well-proven ability to extract underlying features on large and heterogeneous data. Behavioral and linguistic features such as Linguistic Inquiry and Word Count (LIWC), posting time [5], personality, gender, and age [19], topic modeling [22], online engagement and emotion [6], have also been investigated. However, existing depression detection mechanisms harvesting social media suffer from two major limitations. First, existing solutions rely heavily on the amount, quality, and variety

Table 1: Statistics of PsycheNet

	Users	Tweets per user			Friends per user			Tweets per friend			Self-diagnostic tweets	
	Sum	Min	Max	Sum	Min	Max	Sum	Min	Max	Sum	Start	End
Depressed group	372	35	2, 828	357, 662	25	2, 978	78, 112	25	50	3, 831, 547	01/03/2019	07/29/2020
Control group	445	27	3, 186	648, 534	30	879	110, 060	25	50	5, 449, 353	–	–

of content types (textual, visual, etc.) posted by users to make accurate inferences, therefore suffering from the cold-start problem when coping with users with limited social footprints (e.g., most existing works exclude users with fewer than 25 tweets). Second, existing approaches ignore the social impact or indication from users' social circles that can be otherwise utilized to enhance the inference results.

Our Insights on Social Contagion. Besides providing a communication platform, social media enables emotional contagion between users. Recent studies reveal that internet users with stigmatized illnesses such as depression are more likely to use social media for interpersonal communication than people with other chronic illnesses. A study on Facebook revealed that positive and negative emotions can be propagated by online interactions between users, often without their knowledge or awareness [7]. There have been discussions on the link between social media use and increased depression, emptiness, and anxiety in individuals. It was reported that an increase in social media usage can be associated with depressive symptoms such as low self-esteem as users engage in social comparison, envy, and jealousy [1, 29]. Even though the cause of depression is complex, social environments and interpersonal relationships are leading factors that contribute to the development of depression in individuals. Emotion contagion inside social circles has been observed by the medical community [3, 9, 12]. In fact, according to [23], individuals are 93% more likely to be depressed if they are directly connected to a friend who is dealing with depression.

Our Contributions. Motivated by the aforementioned observations, in this paper, we present MentalSpot, a social-contagion based depression early-screening framework using meta-learning. We systematically study an interesting, yet challenging problem: how can we take full advantage of users' social circles (also known as social contagion) to detect the indication of depression as early as possible, i.e., by using as few tweets as possible? This problem is non-trivial due to the following challenges. (1) How can we effectively capture the indication of depression from the linguistic features extracted from limited written texts? (2) How can we distinguish all friends among users' social circles in a way that we can pick out the most predictive friends (i.e., friends who are homogeneous with each target user) to enhance the inference results?

The early detection of depression plays a vital role in supporting healthcare professionals for determining timely and targeted interventions in a clinical setting. Since the inability of existing solutions to support early detection is a serious limitation in practice, our study represents an important step forward for machine learning application in psychiatric diagnosis. To the best of our knowledge, our work is the first of its kind to address the cold-start problem of existing depression detection solutions by using social contagion as

an indicator of an individual's psychological state. We summarize our contributions as follows:

- We present MentalSpot, a social-contagion based framework in support of depression early-screening on social media based on meta-learning, where the implied knowledge of depression at the individuals' level is transferred between friends to combat the data scarcity problem.
- We construct the first social-contagion based data repository, PsycheNet for depression detection, filling the void of social-circle driven depression datasets. Our PsycheNet dataset contains the timelines of Twitter users in depressed and control groups as well as the timelines of the users with whom they are maintaining a bidirectional friendship.
- To effectively capture the indication of depression from the linguistic features implied in limited written texts, we design a triplet network to extract users' linguistic embeddings.
- To distinguish friends who are homogeneous with each target user from her social circles, we employ the dynamic mean shift pruning to identify her top- k homogeneous friends.
- We further train a friend based depression detection model by leveraging the written texts from each target user's top- k homogeneous friends according to the embedding distribution learned from the triplet network in the metric space.
- Extensive experiments show that MentalSpot outperforms the state of the art in terms of all effectiveness metrics, especially for users with very few tweets. Specifically, by using only five tweets per user, MentalSpot successfully yields an F1 score that would otherwise be achieved by the state-of-the-art methods requiring at least twenty tweets.

The remainder of this paper is organized as follows. Section 2 introduces the social-contagion based data repository PsycheNet. Section 3 details the MentalSpot architecture. Comparative experimental evaluations are conducted in Section 4. Section 5 provides the parameter sensitivity analysis. Section 6 presents embedding visualization while Section 7 examines a case study. Finally, related works are reviewed in Section 8 and Section 9 concludes the paper.

2 SOCIAL-CONTAGION DRIVEN DATA REPOSITORY: PSYCHENET

Due to the lack of datasets that investigate depressed users from their social groups, we construct the first social-contagion driven data repository, PsycheNet.¹ Our PsycheNet dataset contains the complete timelines of Twitter users in the depressed and control groups, respectively, as well as the timelines of the users with whom they are maintaining a bidirectional friendship. Table 1 reports the

¹The de-identified dataset may be available upon request and restrictions apply to its use.

Table 2: Samples of sincere and insincere self-diagnostic tweets

Self-diagnostic tweets	
Sincere/Genuine	I've been diagnosed with depression for five years and dealt with it for about ten. Sit down Umm yes! I have been diagnosed with depression in October and this pandemic ain't helping! Ain't nothing wrong with it I been diagnosed with depression I've been diagnosed with depression, and i dont take meds.
Insincere/Fake	Interesting.. you just described me I've been diagnosed with depression but perhaps it's not! Playing bit life and I've been diagnosed with depression at age 3 fuck that I have been diagnosed with depression and depression and depression I've been diagnosed with depression by TikTok

statistics of PsycheNet. Note that since we are motivated to combat the data scarcity problem for the early detection of depression, in order to investigate the performance of MentalSpot in learning from and classifying cold-start users, for any user we only used her m (which defaults to 10) most recent tweets from her self-diagnostic tweet (for the depressed user) or from her last tweet (for the control user) for model training and test in all our experiments.

User Timeline Collection and Ground Truth Verification. We used Twitter Search API to construct PsycheNet. To identify the depressed group, we searched for users who publicly shared their diagnoses by tweeting texts such as “... *I have been diagnosed with depression* ...”. After identifying such individuals, we manually examined and verified the diagnostic tweets for their genuineness before adding them to the dataset. Table 2 shows some sample diagnostic tweets we identified as genuine or fake. For example, we considered as genuine the tweet “*I’ve been diagnosed with depression for five years and dealt with it for about ten.*”. On the other hand, the tweet “*I’ve been diagnosed with depression by TikTok.*” was treated as fake and skipped. For the control group, we randomly selected users who tweeted from the United States and whose tweets did not mention any diagnosis for depression. In addition, the verified Twitter users (accounts with checkmark badges) were skipped since they often have a large number of followers and a relatively low number of followees. We collected up to 3,200 English tweets (excluding retweets) starting from the dates of their self-diagnostic tweets (for the depressed users) or from their last tweets (for the control users) to when their accounts were created. Also, we employed Botometer Pro API [32] to detect users whose activity resembles social bot activity. We removed all the users who scored more than 0.5 on both the Raw and Complete Automation Probability (CAP) metrics.

Collection of Tweets From Social Circles. For each user, we define her “friends” as all Twitter accounts that she follows and also follow her. For each friend of any user in the depressed group, we collected up to 50 English tweets tracing backwards from the date of the self-diagnostic tweet of the depressed user that friend is associated with. On the other hand, for each friend of any user in the control group, we collected up to 50 English tweets tracing backwards from the date of the last tweet of the user that friend is associated with. Also, we excluded any tweets from any friends with a posting date that fell before the account creation date of the user they are associated with.

Tweet Cleaning for Harsh Detection. To not favor the classification of depressed individuals, for any user in the depressed group, we removed any of her tweets that contained the words ‘depression’ or ‘depressed’. On the other hand, we removed from the control group all those users whose tweets contained the words ‘depression’ or ‘depressed’. Furthermore, since the use of first-person singular pronouns is more associated with users with depression than non-depressed individuals [4, 25], we removed all pronouns to minimize the use of domain-specific knowledge towards detecting depression in this work.

3 PROPOSED METHOD

MentalSpot Architecture. Figure 1(a) depicts the architecture of MentalSpot, which includes three modules: (1) user/friend embedding extraction, (2) homogeneous friend selection, and (3) depression detection. To effectively capture the indication of depression from the linguistic features implied in limited written texts, we train an embedding model as part of a triplet network to extract embedding representation for users and their friends. To distinguish friends who are homogeneous with each target user from her social circles, we use dynamic mean shift pruning to identify her homogeneous friends. Then we retrieve her top- k homogeneous friends who are closest to her in the embedding space based on pairwise cosine distance. Subsequently, a friend based depression detection model is trained by leveraging the texts written by each target user’s top- k homogeneous friends identified in the metric space.

User/Friend Embedding Extraction. The embedding model is learned as part of a triplet network [24, 30], an improvement of Siamese network [2], to obtain embedding representations of each user based on a triplet loss over the linguistic features implied in users’ social posts. Figure 1(c) illustrates the architecture of the triplet network in MentalSpot. We divide all users into two categories: control group and depressed group. Each triplet is represented as $(x^a(anchor), x^p(positive), x^n(negative))$, where anchor and positive samples are the aggregated tweets from two users in the same group, while the negative sample is the aggregated tweets from a user in the other group. During training, all the triplet elements are fed to the embedding model which is a convolutional neural network we will introduce soon. The triplet network reduces the triplet loss, defined as follows:

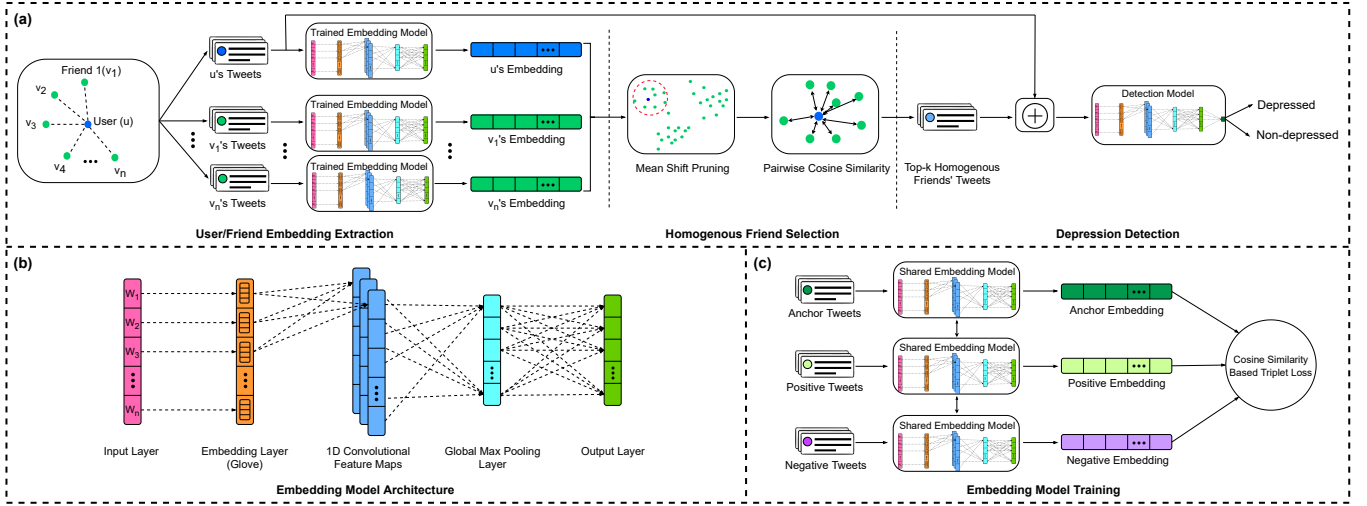


Figure 1: (a) The overall architecture of MentalSpot, where the implied knowledge of depression at the individuals' level is transferred between friends to combat the data scarcity problem; (b) Embedding extraction model; (c) Embedding extraction model pretrained as part of triplet network at the individuals' level.

$$Loss = \sum_{t=1}^N \max(S(\vec{x}_t^a, \vec{x}_t^p) - S(\vec{x}_t^a, \vec{x}_t^p) + m, 0) \quad (1)$$

where $\vec{x}$ is the embedding vector of x , S is the cosine similarity, m represents the constant margin, and N is the cardinality of the set of all triplets. This triplet loss is used to maximize the cosine similarity between the embeddings of each anchor and its positive sample while minimizing the cosine similarity between the embeddings of the anchor and its negative sample. To avoid the selection bias that is associated with hard samples and ensure a fast convergence, we only use semi-hard triplets for training, which are the triplets that meet the following condition:

$$S(\vec{x}^a, \vec{x}^p) > S(\vec{x}^a, \vec{x}^n) > S(\vec{x}^a, \vec{x}^p) - m \quad (2)$$

We treat the embeddings generated by our embedding model as the depression representations of the users/friends, which capture the similarities of their linguistic features in terms of the psychological states. As illustrated in Figure 1(b), we adopt the following architecture and hyper-parameters in the embedding model. We use GloVe word embedding model [13, 18] for initialization. We train a word tokenizer with 100,000 of the most frequent words in our training corpus as the vocabulary set. Subsequently, each input vector, consisting of all concatenated tweets of a user, is tokenized and trimmed to the length of 16,000 (the 90th percentile of input lengths in our training set), which is followed by zero padding. The first layer of our embedding model is an embedding layer with input and output dimensions of 100,000 and 200, respectively. The second layer is a dropout layer with a dropout rate of 0.20. The third layer is a one-dimensional CNN with 200 filters and a kernel size of 3 and stride of 1, followed by a dropout layer with 0.10 as the dropout rate. We adopt a one-dimensional CNN layer rather than

RNN layers [16] since the one-dimensional CNN design yielded the best performance for depression detection in prior work [14]. Next, we apply a global max-pooling layer. Finally, a dense layer with a size of 200 neurons is added to generate the user/friend embeddings.

Homogeneous Friend Selection. For any user/friend, we feed her aggregated tweets as input to the trained embedding model to obtain her embedding. User embeddings are typically high-dimensional and are sparse in the metric space. Therefore, given any target user, we perform the dynamic mean shift pruning in the embedding space to identify her homogeneous friends from her social circles. Using traditional clustering algorithms (i.e., k -means) has two inherent drawbacks: the number of clusters must be known a priori, and the discovered clusters are limited to be spherical or elliptical. On the contrary, as a nonparametric clustering technique, mean shift clustering can overcome these two limitations by iteratively performing kernel density estimation, i.e., estimating the modes (centers of the densest regions) of the multivariate distribution in the feature space. The set of points that converge to the same mode constitutes a cluster. As a result, the number of clusters can be obtained automatically by locating the modes. The original mean shift uses a fixed bandwidth for all the data points. In MentalSpot, we propose to use the **dynamic mean shift** where each user has a different bandwidth based on her friend distribution in the metric space.

Given n data points x_i , $i = 1, \dots, n$, where $x_i \in R^d$, the kernel density estimate can be obtained using

$$f_K(x) = \frac{1}{nb^d} \sum_{i=1}^n K\left(\frac{d(x, x_i)}{b}\right) \quad (3)$$

where $K(x)$ is a symmetric kernel, $d(x, x_i)$ is the distance function between x and x_i in the metric space (cosine distance for overcoming the curse of dimensionality), and b is the bandwidth value. The profile of the kernel $k(x)$ satisfies $K(x) = c_{k,d}k(\|x\|^2)$, where $c_{k,d}$ is a normalization constant that ensures the kernel integrates to 1. The density modes can be obtained by setting the gradient of the density estimator $\nabla f_K(x)$ to 0. By taking the gradient of the density estimator, we have:

$$\nabla f_K(x) = \frac{2c_{k,d}}{nb^{d+2}} \left[\sum_{i=1}^n g \left(\left(\frac{d(x, x_i)}{b} \right)^2 \right) \right] \times \left[\frac{\sum_{i=1}^n x_i g \left(\left(\frac{d(x, x_i)}{b} \right)^2 \right)}{\sum_{i=1}^n g \left(\left(\frac{d(x, x_i)}{b} \right)^2 \right)} - x \right] \quad (4)$$

where $g(x) = -k'(x)$ if the gradient of the kernel profile exists. The mean shift vector can then be written as

$$m_b(x) = \frac{\sum_{i=1}^n x_i g \left(\left(\frac{d(x, x_i)}{b} \right)^2 \right)}{\sum_{i=1}^n g \left(\left(\frac{d(x, x_i)}{b} \right)^2 \right)} - x \quad (5)$$

Note that we compute bandwidth value b for each user based on her friend distribution. In our experiments, we empirically set the bandwidth b for each user to the averaged cosine distance from that user to her first and second closest friends in the embedding space. Given any target user, we employ the proposed dynamic mean shift on all her friends to discover those friends who reside in the same cluster as the target user. We regard those friends who reside in the same cluster as each target user as **her homogeneous friends**. It is worth noting that we also implemented other clustering methods for comparison purpose, including k -means, k -medoids, and agglomerative clustering, but none of them outperformed the proposed dynamic mean shift method. Subsequently, we retrieve each target user's top- k closest homogeneous friends. The top- k closest homogeneous friends of user u is the set of u 's friends who are closest to u in the embedding space and also in the same cluster as u . Algorithm 1 describes how MentalSpot discovers the top- k homogeneous friends for each user.

Depression Detection. After identifying the top- k homogeneous friends for each target user, we gather the tweets from those friends and concatenate them with the tweets of each target user they are associated with. Here our goal is to incorporate the tweet correspondence between friends to transfer the implied knowledge of depression captured by the triplet network at the individuals' level to combat the data scarcity problem [10, 27]. Similar to [14], we treat each user's timeline as a document. After tokenizing the concatenated tweets, we create a unique input vector for each target user in the depressed and control groups by using GloVe word embedding. The depression detection model follows the same architecture and hyper-parameters as used in the embedding model except that it has a binary classification head, as shown in Figure 1(a).

Algorithm 1: Top- k Homogeneous Friend Discovery

Input: User set $\mathcal{U}$, user tweet set $\mathcal{T}$, friend set $\mathcal{U}'$, friend tweet set $\mathcal{T}'$, k

Output: Top- k homogeneous friends for each user u in $\mathcal{U}$

```

1 for Each user  $u$  in  $\mathcal{U}$  do
2   Retrieve the embedding  $\vec{u}$  of  $u$  by feeding  $u$ 's tweets in
    $\mathcal{T}$  to the pre-trained triplet embedding model.
3   for Each friend  $f$  of  $u$  in  $\mathcal{T}'$  do
4     Retrieve the embedding  $\vec{f}$  of  $f$  by feeding  $f$ 's tweets
     in  $\mathcal{T}'$  to the pre-trained triplet embedding model.
5   Compute the bandwidth value for  $u$ .
6   Apply the dynamic mean shift to  $\vec{u}$  and all friends of  $u$ 
   in  $\mathcal{U}'$ .
7   for Each friend  $f$  of  $u$  assigned in the same cluster as  $u$  do
8     similarity = CosineSimilarity( $\vec{u}$ ,  $\vec{f}$ )
9   Sort all friends of  $u$  assigned in the same cluster as  $u$  by
   the similarity score.
10  Retrieve the  $k$  friends of  $u$  with the highest similarity
   scores.
```

4 EXPERIMENTAL VALIDATION

In this section, we empirically show the superiority of MentalSpot over the state of the art. To ensure that there was no data leakage during model training and test, we first partitioned all Twitter users in PsycheNet into training and test sets with a split of 80% and 20%. Second, we removed any overlap between the friends of those users in the training set and the friends of those users in the test set. Table 3 reports the experimental parameters along with their default values used in this section.

Table 3: Experimental parameters. The asterisk (*) denotes the default parameter value

Parameter	Symbol	Values			
Number of tweets per user	m	5	10*	15	20
Number of top friends per user	k	1	3	5*	7
Number of tweets per top friend	n	5	10*	15	20

4.1 Methods of Comparison

We implemented the following methods for comparison.

- **TF-IDF+SVM** [21]. This method employs TF-IDF to extract features from a user's tweets. The features are then passed to SVM with a linear kernel for classification.
- **CNNWithMax** [14]. This method passes each user's concatenated tweets to an embedding layer, followed by a one-dimensional CNN architecture.
- **CNN+fastText** [28]. This method feeds the user's concatenated tweets to a two-dimensional CNN architecture for detecting depression, based on fastText embedding.
- **LSTM+fastText** [31]. This method passes a user's concatenated tweets to three stacked Long Short Term Memory

(LSTM) layers to preserve sequential information for detecting depression, based on fastText embedding.

- **MentalSpot**. This is the complete version of our proposed MentalSpot method based on social contagion.
- **MentalSpot-Rand**. This is a degraded version of MentalSpot with friends randomly selected from users' social circles.

4.2 Performance Comparison

We investigated the performance of all the models for detecting depression in users by conducting five-fold cross-validation and reported F1 score, recall, precision as the effectiveness metrics for evaluation.

Impact on F1 Score: Figure 2 shows the impact of the number, m , of tweets we use per user on the model performance in terms of F1 score. Specifically, Figure 2(a) depicts the F1 scores for the depressed group while Figure 2(b) illustrates the F1 scores for the control group. As shown in Figure 2(a), MentalSpot significantly and consistently outperformed the other models in classifying users into the depressed group, especially when user tweets were limited. For instance, when m was 5, MentalSpot, MentalSpot-Rand, CNN+fastText, CNNWithMax, LSTM+fastText, and TF-IDF+SVM yielded F1 scores of 0.76, 0.68, 0.66, 0.68, 0.69, 0.63, respectively. When m was 20, MentalSpot, MentalSpot-Rand, CNN+fastText, CNNWithMax, LSTM+fastText, and TF-IDF+SVM returned F1 scores of 0.85, 0.75, 0.77, 0.80, 0.73, 0.74, respectively. With respect to the control group, as shown in Figure 2(b), MentalSpot also outperformed all the other models. For instance, when m was set at 10, MentalSpot, MentalSpot-Rand, CNN+fastText, CNNWithMax, LSTM+fastText, and TF-IDF+SVM returned F1 scores of 0.84, 0.80, 0.79, 0.79, 0.76, 0.79, respectively. The superiority of MentalSpot can be explained by its taking advantage of the social footprints left by their friends.

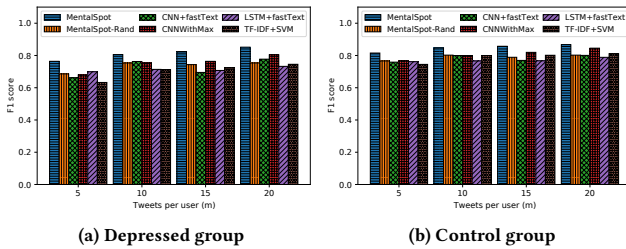


Figure 2: Performance comparison of F1 score

Impact on Recall: Figure 3 displays the impact of the number, m , of tweets we use per user on the model performance in terms of recall score. Specifically, Figure 3(a) depicts the recall score comparison for the depressed group while Figure 3(b) illustrates the recall score comparison for the control group. As shown in Figure 3(a), MentalSpot significantly outperformed all other models. For example, when m was set at 5, MentalSpot, MentalSpot-Rand, CNN+fastText, CNNWithMax, LSTM+fastText, and TF-IDF+SVM yielded recall scores of 0.73, 0.64, 0.60, 0.62, 0.68, 0.56, respectively. On the other hand, as shown in Figure 3(b), MentalSpot outperformed all other models for the control group when m was less than 20.

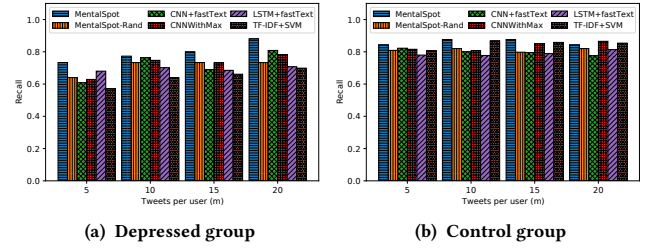


Figure 3: Performance comparison of recall

Impact on Precision: Figure 4 shows the impact of the number, m , of tweets we use per user on the model performance in terms of precision score. Specifically, Figure 4(a) depicts the precision scores for the depressed group while Figure 4(b) displays the precision scores for the control group. As shown in Figure 4(a), MentalSpot consistently outperformed all other models for the depressed group. For example, with m at 5, MentalSpot, MentalSpot-Rand, CNN+fastText, CNNWithMax, LSTM+fastText, and TF-IDF+SVM yielded precision scores of 0.79, 0.73, 0.78, 0.74, 0.72, 0.71, respectively. As shown in Figure 4(b), MentalSpot also consistently outperformed all the other models for the control group. For example, with m at 10, MentalSpot, MentalSpot-Rand, CNN+fastText, CNNWithMax, LSTM+fastText, and TF-IDF+SVM achieved precision scores of 0.82, 0.78, 0.80, 0.79, 0.75, 0.74, respectively.

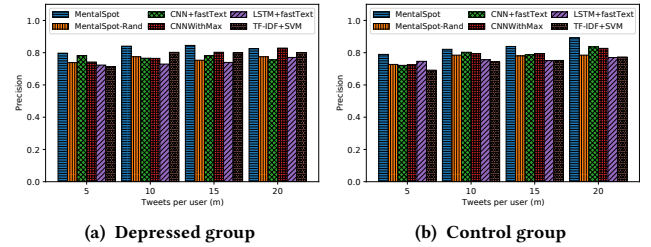


Figure 4: Performance comparison of precision

5 PARAMETER SENSITIVITY ANALYSIS

5.1 Impact of the Number of Top Friends Per User

We investigated the performance of MentalSpot by varying the number, k , of the top friends we use per user. We set the number, m , of tweets per user and the number, n , of tweets per each friend to their default values of 10. We conducted five-fold cross-validation and reported F1 score as the effectiveness metric for evaluation. Figure 5 shows the impact of k on the performance of MentalSpot and MentalSpot-Rand in terms of F1 score. Specifically, Figure 5(a) depicts the F1 scores for the depressed group while Figure 5(b) illustrates the F1 scores for the control group. As shown in Figure 5(a), when k was set at 1, 3, 5, and 7, MentalSpot yielded F1 scores of 0.79, 0.77, 0.80, 0.79 while MentalSpot-Rand yielded F1 scores of 0.75, 0.72, 0.75, 0.76. As shown in Figure 5(b), as the value of k increased from 1 to 7, MentalSpot returned F1 scores of 0.80, 0.81,

0.84, 0.82 while MentalSpot-Rand returned F1 scores of 0.78, 0.75, 0.80, 0.81. We can observe that increasing the number of friends per each user from 1 to 5 led to better performance in MentalSpot. However, when k was set at 7, the performance of MentalSpot did not improve even though the training records were increased. This verifies the need of selecting homogeneous friends since the top-5 friends retrieved in MentalSpot were the most homogeneous friends and therefore, simply increasing the number of friends will not necessarily improve the performance, let alone using all friends.

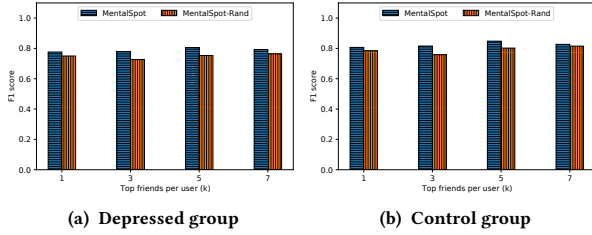


Figure 5: Impact of k on F1 score of MentalSpot

5.2 Impact of the Number of Tweets Per Top Friend

We also investigated the performance of MentalSpot by varying the number, n , of tweets we use per top friend. We set the number, m , of tweets per user and the number, k , of the top friends per user to their default values of 10 and 5. Figure 6 shows the impact of n on the model performance in terms of F1 score. Specifically, Figure 6(a) depicts the F1 scores for the depressed group, while Figure 6(b) reports the F1 scores for the control group. As shown in Figure 6(a), when n was set at 5, 10, 15, and 20, MentalSpot yielded F1 scores of 0.78, 0.80, 0.80, 0.80 while MentalSpot-Rand yielded F1 scores of 0.70, 0.75, 0.77, 0.74. As shown in Figure 6(b), MentalSpot returned F1 scores of 0.82, 0.84, 0.83, 0.84 while MentalSpot-Rand returned F1 scores of 0.73, 0.80, 0.81, 0.80, as the value of n increased from 5 to 20. We can observe that when we increased the number of tweets from 5 to 10 per friend, the performance improved. When we kept increasing the number of tweets per friend, the performance gain was limited.

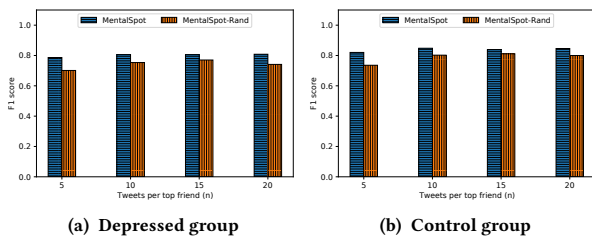


Figure 6: Impact of n on F1 score of MentalSpot

6 USER EMBEDDING AUGMENTED WITH FRIEND TWEETS

To further illustrate the effectiveness of using friend tweets for enhancing the performance of depression detection, we show the 2d-visualization of the user embeddings generated by the depression detection module (i.e., the input to the final output layer in the depression detection module of MentalSpot) for the control and depressed users in the test set. Figure 7 shows the user embeddings learned using 5 tweets for each target user with and w/o tweets from their friends while Figure 8 shows the user embeddings learned using 20 tweets for each target user with and without tweets from their friends. By comparing Figure 7(a) and Figure 7(b), we can observe that the test users in the depressed and control groups became more separated when using friend tweets than not using friend tweets, given 5 tweets per test user. Similarly, as shown in Figure 8(a) and Figure 8(b), the test users in the depressed and control groups became more distinguishable when using friend tweets than not using friend tweets, given 20 tweets per test user.

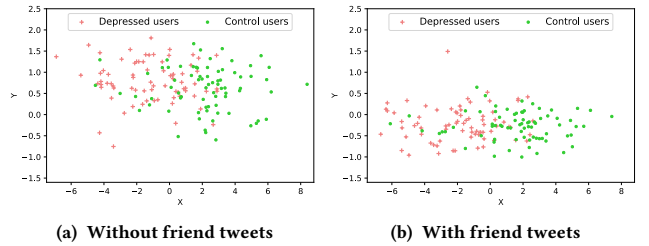


Figure 7: Comparison of user embeddings learned with and w/o friend tweets. We used 5 tweets per user and 10 tweets per the top-5 friend. The red crosses represent the test users in the depressed group while the green dots represent the test users in the control group.

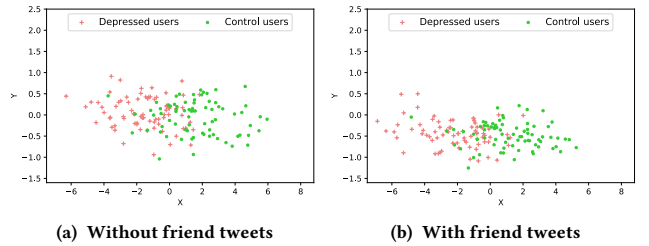


Figure 8: Comparison of user embeddings learned with and w/o friend tweets. We used 20 tweets per user and 10 tweets per the top-5 friend. The red crosses represent the test users in the depressed group while the green dots represent the test users in the control group.

7 CASE STUDY

We randomly selected two target users (one from depressed group and the other from control group) from the test set to investigate their relationships with their respective friends by using the embeddings learned in MentalSpot. We first visualize the distributions

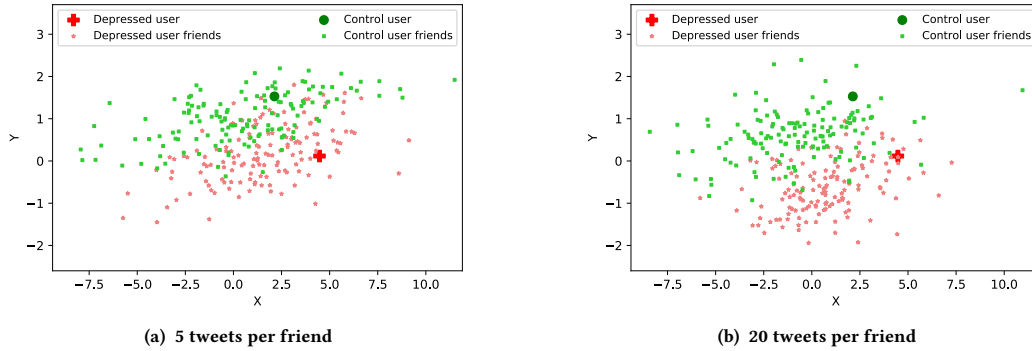


Figure 9: Embedding distributions of two users randomly selected from test set and their friends. The big red cross represents the sample user from the depressed group while the big green dot represents the sample user from the control group.

of their friends’ embeddings, followed by a linguistic comparison of them with their most and least similar friends in the metric space.

7.1 Friend Distribution of Sample Test Users

Figure 9 depicts the embedding representations for two users randomly selected from the test set, along with those of their respective friends. The big red cross represents the embedding of the target user randomly selected from the depressed group while the big green dot means the embedding of the target user randomly selected from the control group. Here we used 20 tweets per user in the training set to train the embedding model, which was further used to generate the embeddings for all the users in the test set as well as their corresponding friends. We varied the number of tweets used in each friend to investigate its impact on their embeddings. Specifically, Figure 9(a) plots the friends’ embeddings using 5 tweets per friend while Figure 9(b) depicts the friends’ embeddings using 20 tweets per friend. A few interesting observations can be made here. First, two target users’ embeddings are surrounded by their friends’ embeddings, which verifies our hypothesis that the linguistic features exhibited by friends can be exploited to indicate the depression state of each user. Second, when we increased the number of tweets used for each friend, the social circles of the depressed user and the control user became more distinguishable in the metric space, leading to a decrease in the overlap between the friends’ clusters. Third, with an increase in the number of tweets used for each friend, all the friends of a target user (either a depressed user or a control user) become more separable, which verifies the effectiveness of our embedding model to capture the similarities between the target users and their homogeneous friends as well as the dissimilarities between the target users and their heterogeneous friends.

7.2 Linguistic Comparison of Homogeneous Friends

For the two target users we selected in the previous subsection, we identified their most and least similar friends in the metric space based on the embeddings generated by MentalSpot. We plotted the word clouds of the two target users and their corresponding most and least similar friends. Notice that the embeddings generated by

MentalSpot aim to reflect the level of existence of depression in individuals, rather than simply the occurrence frequency of words used by each user. However, showing the most frequently used words (by using word clouds) can provide us with more insight into how the vocabularies used differ among users.

Figure 10 illustrates the top 50 words derived from the recent 20 tweets of the depressed target user along with her most and least homogeneous friends. We can observe that the words appear in Figure 10(a) have more in common with the words shown in Figure 10(b) than those in Figure 10(c). Specifically, in Figure 10(a) and Figure 10(b), there are words that are usually more associated with depression, such as *suicide*, *upset*, *lose*, *die*, *hate*, *dead*, and *sad*, which do not occur in Figure 10(c). On the other hand, Figure 11 illustrates the top 50 words derived from the recent 20 tweets of the control target user along with her most and least homogeneous friends. We can observe that Figure 11(c) includes words that are more associated with negative feelings, such as *sad*, *frustrated*, *frustrating*, *anxiety*, *inability*, and *fail*. Such comparison of linguistic features between target users and their most and least homogeneous friends demonstrates the effectiveness of the homogeneous friend selection in MentalSpot.

8 RELATED WORK

Machine Learning Based Depression Detection. Statistical and traditional machine learning-based approaches rely heavily on handcrafting features which is a time-consuming process and also requires specific domain knowledge. Choudhury et al. [6] investigated whether social media can be used to detect and predict major depressive disorder in users. Features such as depressive words, mentions of antidepressant medications, emotional states, and dominance were considered. In addition, user behavioral features were taken into account by extracting social engagement and ego-network features. The dimensionality of the features was reduced using Principal Component Analysis (PCA) and then the vectors were fed to a binary classifier constructed by SVM. Preotiuc-Pietro et al. [19] explored the potential link between personality types and two of the mental health disorders: depression and post-traumatic stress disorder. Features such as age, personality, gender,



Figure 10: Word clouds of the random depressed user and her most and least homogeneous friends

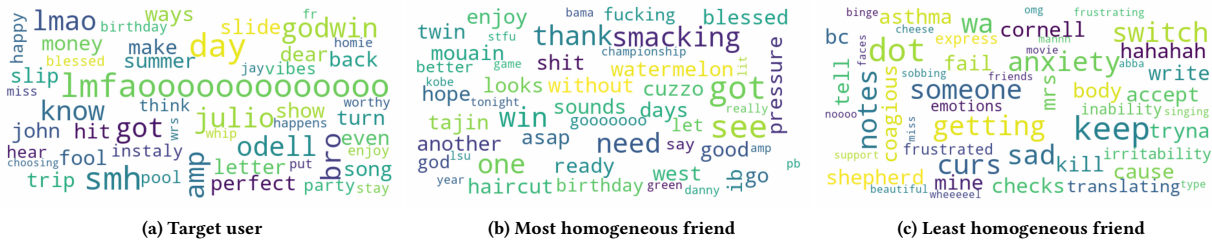


Figure 11: Word clouds of the random control user and her most and least homogeneous friends

affect, and intensity along with linguistic features were considered. By considering the Patient Health Questionnaire-9 (PHQ-9), Yazdavar et al. [34] built a depression lexicon based on Twitter users who shared their symptoms. Sentiment analysis was performed in order to build a seeded Latent Dirichlet Allocation (LDA) capable of monitoring clinical depression symptoms. Shen et al. [26] considered clinical depression criteria to derive domain-specific features and online behavior of Twitter users to extract LDA topics, social network, visual, user profile, and emotional features. The combined depressive features were used in a dictionary learning model with logistic regression to detect depression for Twitter users.

Deep Learning Based Depression Detection. In contrast to statistical and traditional approaches, deep-learning based approaches do not rely heavily on hand-crafted features. Over the last decade, deep learning techniques have gained popularity in various natural language processing tasks such as sentiment analysis, spam detection, and more importantly, mental health detection. Yates et al. [33] utilized CNN architecture to extract vector representations of users' activities on Reddit. Orabi et al. [14] employed CBOW and skip-gram models to capture linguistic features for Twitter users and all users' tweets were then concatenated and fed to three different CNN models along with a BiLSTM with context-aware attention model for depression detection. Gui et al. [11] utilized a reinforcement learning approach to train an agent capable of identifying depressed users based on textual and visual features on Twitter. In their design, Gated Recurrent Units (GRUs) were employed to extract textual features while the pre-trained VGG-16 model was utilized for the visual feature extraction. Zogan et al. [35] introduced a hybrid model to consider both behavioral and linguistic features for detecting depression for Twitter users. In

their design, a GRU-based architecture was adopted to learn from topical, temporal, and semantic features while a CNN model was used for extracting linguistic features from users' tweets. Unfortunately, all the aforementioned depression detection mechanisms suffer from two major limitations. First, they rely heavily on the amount, quality, and variety of content types posted by users to make accurate inferences, therefore suffering from the cold-start problem when coping with users with limited social footprints. Second, they fail to consider the social impact or indication that can be taken advantage of from users' social circles to enhance the inference results for early depression detection.

9 CONCLUSION AND FUTURE WORK

In this paper, we present MentalSpot, a social contagion-based depression early-screening framework based on meta-learning, where the implied knowledge of depression at the individuals' level is transferred between friends to combat the data scarcity problem. The early detection of depression plays a vital role in supporting healthcare professionals for determining timely and targeted interventions in a clinical setting. Since the inability of existing solutions to support early detection is a serious limitation in practice, our study represents an important step forward for machine learning application in psychiatric diagnosis. To the best of our knowledge, our work is the first of its kind to address the cold-start problem of existing depression detection solutions by using social contagion. Extensive experiments show that MentalSpot outperforms the state of the art in terms of all effectiveness metrics, especially for users with very few tweets. Our future agenda includes the extension of this framework to other psychiatric diagnoses and prospective model deployment for the prediction of complications.

REFERENCES

- [1] Helmut Appel, Jan Crusius, and Alexander L Gerlach. 2015. Social comparison, envy, and depression on Facebook: A study looking at the effects of high comparison standards on depressed individuals. *Journal of Social and Clinical Psychology* 34, 4 (2015), 277–289.
- [2] Jane Bromley, James W Bentz, Léon Bottou, Isabelle Guyon, Yann LeCun, Cliff Moore, Eduard Säckinger, and Roopak Shah. 1993. Signature verification using a “siamese” time delay neural network. *International Journal of Pattern Recognition and Artificial Intelligence* 7, 04 (1993), 669–688.
- [3] John T Cacioppo, James H Fowler, and Nicholas A Christakis. 2009. Alone in the crowd: the structure and spread of loneliness in a large social network. *Journal of personality and social psychology* 97, 6 (2009), 977.
- [4] Glen Coppersmith, Mark Dredze, and Craig Harman. 2014. Quantifying Mental Health Signals in Twitter. In *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*. Association for Computational Linguistics, Baltimore, Maryland, USA, 51–60. <https://doi.org/10.3115/v1/W14-3207>
- [5] Munmun De Choudhury, Scott Counts, and Eric Horvitz. 2013. Social Media as a Measurement Tool of Depression in Populations. In *Proceedings of the 5th Annual ACM Web Science Conference (Paris, France) (WebSci '13)*. Association for Computing Machinery, New York, NY, USA, 47–56. <https://doi.org/10.1145/2464464.2464480>
- [6] Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. 2013. Predicting depression via social media. *Icwm* 13 (2013), 1–10.
- [7] Michela Del Vicario, Gianna Vivaldo, Alessandro Bessi, Fabiana Zollo, Antonio Scala, Guido Caldarelli, and Walter Quattrociocchi. 2016. Echo chambers: Emotional contagion and group polarization on facebook. *Scientific reports* 6 (2016), 37825.
- [8] Centers for Disease Control and Prevention. 2017. WISQARS Leading Causes of Death Reports. <https://webappa.cdc.gov/sasweb/ncipc/leadcause.html>. Accessed: 2020-11-08.
- [9] James Fowler and Nicholas Christakis. 2008. Dynamic Spread of Happiness in a Large Social Network: Longitudinal Analysis Over 20 Years in the Framingham Heart Study. *BMJ (Clinical research ed.)* 337 (02 2008), a2338. <https://doi.org/10.1136/bmj.a2338>
- [10] Madhuri Ghorpade, Haiquan Chen, Yuhong Liu, and Zhe Jiang. 2020. SMART: Emerging Activity Recognition with Limited Data for Multi-modal Wearable Sensing. In *2020 IEEE International Conference on Big Data (Big Data)*. 1316–1321. <https://doi.org/10.1109/BigData50022.2020.9377746>
- [11] Tao Gui, Liang Zhu, Qi Zhang, Minlong Peng, Xu Zhou, Keyu Ding, and Zhigang Chen. 2019. Cooperative Multimodal Approach to Depression Detection in Twitter. *Proceedings of the AAAI Conference on Artificial Intelligence* 33, 01 (Jul. 2019), 110–117. <https://doi.org/10.1609/aaai.v33i01.3301110>
- [12] Jamie Guillory, Jason Spiegel, Molly Drislane, Benjamin Weiss, Walter Donner, and Jeffrey Hancock. 2011. *Upset Now? Emotion Contagion in Distributed Groups*. Association for Computing Machinery, New York, NY, USA, 745–748. <https://doi.org/10.1145/1978942.1979049>
- [13] Bo Hui, Haiquan Chen, Da Yan, and Wei-Shinn Ku. 2021. EDGE: Entity-Diffusion Gaussian Ensemble for Interpretable Tweet Geolocation Prediction. In *2021 IEEE 37th International Conference on Data Engineering (ICDE)*. 1092–1103. <https://doi.org/10.1109/ICDE51399.2021.00099>
- [14] Ahmed Hussein Orabi, Prasadith Buddhitha, Mahmoud Hussein 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*. Association for Computational Linguistics, New Orleans, LA, 88–97. <https://doi.org/10.18653/v1/W18-0609>
- [15] Md Rafiqul Islam, Abu Kamal, Naznin Sultana, Robiul Islam, Anwaar Ulhaq, and Mohammad Moni. 2018. Detecting Depression Using K-Nearest Neighbors (KNN) Classification Technique. In *2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2)*. IEEE, Bangladesh, 1–4. <https://doi.org/10.1109/IC4ME2.2018.8465641>
- [16] Avik Nayak, Haiquan Chen, Xiaojun Ruan, and Jinsong Ouyang. 2019. DeepSpot: Understanding Online Opinion Spam by Text Augmentation using Sentiment Encoder-Decoder Networks. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Analytics for Local Events and News, November 5, 2019, Chicago, Illinois, USA*. ACM, 3:1–3:10. <https://doi.org/10.1145/3356473.3365187>
- [17] World Health Organization et al. 2017. *Depression and other common mental disorders: global health estimates*. Technical Report. World Health Organization.
- [18] Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. GloVe: Global Vectors for Word Representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, Doha, Qatar, 1532–1543. <https://doi.org/10.3115/v1/D14-1162>
- [19] Daniel Preotiuc-Pietro, Johannes Eichstaedt, Gregory Park, Maarten Sap, Laura Smith, Victoria Tobolsky, H. Andrew Schwartz, and Lyle Ungar. 2015. The role of personality, age, and gender in tweeting about mental illness. In *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*. Association for Computational Linguistics, Denver, Colorado, 21–30. <https://doi.org/10.3115/v1/W15-1203>
- [20] 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), 1–11.
- [21] Philip Resnik, William Armstrong, Leonardo Claudino, and Thang Nguyen. 2015. The University of Maryland CLPsych 2015 Shared Task System. In *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*. Association for Computational Linguistics, Denver, Colorado, 54–60. <https://doi.org/10.3115/v1/W15-1207>
- [22] Philip Resnik, William Armstrong, Leonardo Claudino, Thang Nguyen, Viet-An Nguyen, and Jordan Boyd-Graber. 2015. Beyond LDA: Exploring Supervised Topic Modeling for Depression-Related Language in Twitter. In *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*. Association for Computational Linguistics, Denver, Colorado, 99–107. <https://doi.org/10.3115/v1/W15-1212>
- [23] J Niels Rosenquist, James H Fowler, and Nicholas A Christakis. 2011. Social network determinants of depression. *Molecular psychiatry* 16, 3 (2011), 273–281.
- [24] Florian Schroff, Dmitry Kalenichenko, and James Philbin. 2015. FaceNet: A unified embedding for face recognition and clustering. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Boston, MA, USA, 815–823. <https://doi.org/10.1109/CVPR.2015.7298682>
- [25] H. Andrew Schwartz, Johannes Eichstaedt, Margaret L. 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*. Association for Computational Linguistics, Baltimore, Maryland, USA, 118–125. <https://doi.org/10.3115/v1/W14-3214>
- [26] Guangyao Shen, Jia Jia, Liqiang Nie, Fuli Feng, Cunjun Zhang, Tianrui Hu, Tat-Seng Chua, and Wenwu Zhu. 2017. Depression Detection via Harvesting Social Media: A Multimodal Dictionary Learning Solution. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*. ijcai.org, Melbourne, Australia, 3838–3844. <https://doi.org/10.24963/ijcai.2017/536>
- [27] Ting Shen, Haiquan Chen, and Wei-Shinn Ku. 2018. Time-aware location sequence recommendation for cold-start mobile users. In *Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, SIGSPATIAL 2018, Seattle, WA, USA, November 06-09, 2018*. ACM, 484–487. <https://doi.org/10.1145/3274895.3274958>
- [28] Marcel Trotzek, Sven Koitka, and Christoph Friedrich. 2018. Utilizing Neural Networks and Linguistic Metadata for Early Detection of Depression Indications in Text Sequences. *IEEE Transactions on Knowledge and Data Engineering* 32 (04 2018), 588–601. <https://doi.org/10.1109/TKDE.2018.2885515>
- [29] Erin A Vogel, Jason P Rose, Lindsay R Roberts, and Kathryn Eckles. 2014. Social comparison, social media, and self-esteem. *Psychology of Popular Media Culture* 3, 4 (2014), 206.
- [30] Jiang Wang, Yang Song, Thomas Leung, Chuck Rosenberg, Jingbin Wang, James Philbin, Bo Chen, and Ying Wu. 2014. Learning Fine-Grained Image Similarity with Deep Ranking. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, Columbus, OH, USA, 1386–1393. <https://doi.org/10.1109/CVPR.2014.180>
- [31] JT Wolohan. 2020. Estimating the effect of COVID-19 on mental health: Linguistic indicators of depression during a global pandemic. In *Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020*. Association for Computational Linguistics, Online. <https://www.aclweb.org/anthology/2020.nlpCOVID19-acl.12>
- [32] Kai-Cheng Yang, Onur Varol, Clayton A Davis, Emilio Ferrara, Alessandro Flammini, and Filippo Menczer. 2019. Arming the public with artificial intelligence to counter social bots. *Human Behavior and Emerging Technologies* 1, 1 (2019), 48–61.
- [33] Andrew Yates, Arman Cohan, and Nazli Goharian. 2017. Depression and Self-Harm Risk Assessment in Online Forums. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Copenhagen, Denmark, 2968–2978. <https://doi.org/10.18653/v1/D17-1322>
- [34] Amir Hossein Yazdavar, Hussein S. Al-Olimat, Monireh Ebrahimi, Goonmeet Bajaj, Tanvi Banerjee, Krishnaprasad Thirunarayan, Jyotishman Pathak, and Amit Sheth. 2017. Semi-Supervised Approach to Monitoring Clinical Depressive Symptoms in Social Media. In *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017 (Sydney, Australia) (ASONAM '17)*. Association for Computing Machinery, New York, NY, USA, 1191–1198. <https://doi.org/10.1145/3110025.3123028>
- [35] Hamad Zogan, Xianzhi Wang, Shoaib Jameel, and Guandong Xu. 2020. Depression Detection with Multi-Modalities Using a Hybrid Deep Learning Model on Social Media. *CoRR abs/2007.02847* (2020). arXiv:2007.02847 <https://arxiv.org/abs/2007.02847>