Knowledge Augmentation for Early Depression Detection

Hrishikesh Kulkarni, Sean MacAvaney, Nazli Goharian and Ophir Frieder

Abstract Individuals continue to share their mental health concerns on social media, providing an avenue to rapidly detect those potentially in need of assistance. While users of immediate need can be recognized with relative ease, early-stage disorder users in the boundary region pose a greater challenge to detect. The minimal posting histories of such users further complicate proceedings. However, these same boundary region users would benefit greatly from timely treatment; hence, detecting their mental health status is of utmost need. Additionally, pointers to identify the type of depression could be of great help. Augmenting knowledge for low posting users can help to solve this problem. We propose an NLP based method 'STBound' that intelligently determines the optimal region for knowledge augmentation. It answers three crucial questions: when?, for whom? and how much? to augment - to resolve this imbroglio. Our proposed selective knowledge augmentation method contributes to early depression detection performance improvement by an average of 11.9% in F1 score. Further, this approach shows promising performance enhancement of 12.1% in F1 score for the critical task of separating these boundary region users with bipolar depression. STBound identifies those depressed users in the boundary region who would otherwise go unidentified.

Key words: depression detection, social media, mental health, early risk detection

Hrishikesh Kulkarni
Georgetown University, Washington DC, USA e-mail: hpk8@georgetown.edu

Sean MacAvaney
University of Glasgow, Glasgow, UK e-mail: first.last@glasgow.ac.uk

Nazli Goharian
IR Lab, Georgetown University, Washington DC, USA e-mail: first@ir.cs.georgetown.edu

Ophir Frieder
IR Lab, Georgetown University, Washington DC, USA e-mail: first@ir.cs.georgetown.edu

1 Introduction

Depression, a mental state resulting in lack of hope and dejection along with persistent sadness [36], is a global health concern. Severe depression may result in self-harm, including suicide, and is also a major cause of disability worldwide. Around 5% [36] of the world population is estimated to be suffering from depression with 80% of those afflicted between 16 and 65 years of age, namely the working population. Thus, in addition to a social health issue, depression directly impacts the overall economy [20]. For at least the aforementioned reasons, early depression detection is of utmost importance. Unfortunately, the criticality to address depression only further increased recently due to reduced social interactions and lifestyle changes caused by the COVID-19 pandemic [13] resulting in a 25% increase in anxiety and depression [6].

A significant percentage i.e., around 16 to 20% of total depressed individuals qualify for bipolar disorder, and the median age of onset for bipolar disorder is 25 years [14]. Considering the high risk of unresolved morbidity associated with bipolar disorders, there is a pressing need to separate bipolar disorders in early stages.

Social media provides users venues to express their thoughts and feelings in the form of written posts. As mental health is a key factor leading to changes in textual patterns and articulations [34], noting these changes can contribute to identifying potential depression [39]. Unfortunately, existing depression detection systems generally require a substantial volume of data to predict [37, 39], resulting in depression being detected only at a later stage, potentially increasing disease severity. Complicating detection, many users have low posting activity, either due to their low posting frequency or simply being new to a platform. We refer to these users as low-resource users and focus our attention towards them, believing that early detection, irrespective of the number of posts, can help a larger population segment. We further empirically and mathematically define low-resource users with extensive analysis on Reddit Self-reported Depression Diagnosis (RSDD) dataset [37]. Usually, bipolar depression in case of low-resource users cannot be differentiated from major depression [21]. This results in sub-optimal treatment and poor outcome in this case [18]. The treatment for bipolar depression is significantly different. Hence, separating it at an early stage in such low-resource users could prove to be crucial.

Broadly speaking, our focus is on identifying low-resource users on the brink, namely those users in the boundary region of depression with a low posting frequency. We propose an NLP based method to intelligently identify low-resource social media users, potentially suffering from depression. The proposed method focuses on early depression detection and provides pointers to separate users with bipolar disorder. Depression detection is based on textual social media posts. These linguistic expressions by users are used to detect their mental state. Although, as described later, others have focused on similar detection, our primary attention is on those hard-to-detect users with low number of postings.

Our contributions are as follows:

- We identify low-resource users in need of help with detailed empirical and mathematical analysis and establish a lower bound on the δ parameter for correct re-evaluation of depressed users in the boundary region.
- We develop an approach to identify low-resource, at-risk, boundary users on the brink of depression.
- We demonstrate that our proposed intelligent and selective knowledge augmentation significantly increases early depression detection accuracy.

2 Related Work

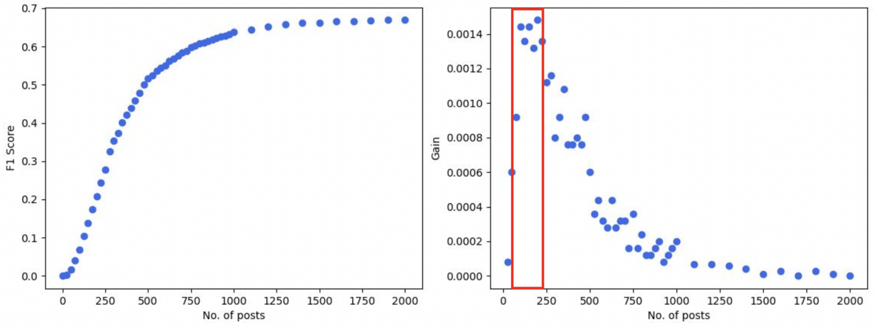
Increasing social media use has created additional venues for continuous mental health monitoring [30, 11, 37, 27, 12, 7]. A number of forums exit that help users with mental health problems via counselling by moderators. Identifying users with such immediate need is crucial in this process. Thus, triaging with high accuracy is necessary for prioritizing users to seek timely help [8, 9, 1]. Social media posts provide timely linguistic cues for mental health monitoring [34]. Researchers also worked on identification of linguistic cues for depression detection based on lexicons [35]. Depression is evident from social media behavior which comprises of use of language over time and sequence of posts [10]. Interestingly, language use itself is a prominent indicator of depression [22]. Typically, neural network based methods deliver better performance in identifying depressed users based on language usage [37].

RSDD dataset is an extensive self-reported depression diagnosis dataset constructed from Reddit [37]. Considering the relevance of timestamps of postings and resulting dynamic behavior, temporal cues were identified on the RSDD dataset [27]. While the RSDD dataset contains depression labels, Self-Reported Mental Health Diagnoses (SMHD) is a comprehensive dataset which provides self-reported diagnosis for bipolar and major depression along with other mental health conditions [7]. It contains Reddit posts of large set of control users, and a few thousand bipolar and depressed users. Dataset with diverse-mental health conditions can provide understanding into mental health related language which can be further leveraged to obtain crucial insights into mental health conditions [7]. Researchers also worked on extraction of medical concepts from large-scale datasets [31]. Datasets with mental health posts and corresponding human-written summaries have been constructed to facilitate mental health research [32]. RSDD and SMHD datasets have been widely used for performance evaluation of different methods.

Bipolar Disorder (BD) is the 10^{th} most common cause of frailty in young population [33] and has triggered serious consequences. It affects life expectancy by 9 to 17 years [33]. It is a mental disorder with high prevalence, but can be misdiagnosed as a major depressive disorder [29]. 40% of patients with bipolar disorder are first misdiagnosed as major depressive disorder [33] and 17% of patients diagnosed as major depressive disorder were found to have undiagnosed bipolar disorder [15]. This makes it exceedingly important to separate users with bipolar disorder. Researchers worked on detecting bipolar disorders using neural network and radial basis function [25]. Different Machine Learning techniques like Decision Trees, Random Forest, SVM, Naïve Bayes, Logistic Regression and KNN were tried out to separate users with bipolar disorder [2].

On the other hand, it is also important to detect depression early to provide timely help before situation slips out of control. Apart from a major focus on a large number of posts, few researchers worked on early depression detection [3, 5, 9]. Researchers also proposed neural models to simplify medical text for consumption of general users using medical social media text [28]. Transformer based techniques were explored for applications in various fields like mental health [17]. Effectively adjusting sensitivity of classifiers can contribute to significant performance leap irrespective of classification methods [23].

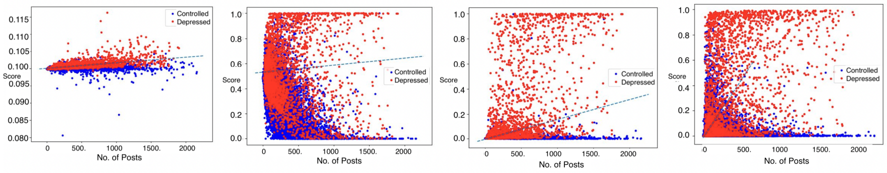
Attempts thus far use large volumes of data. Limited exploration has been carried out on detection of depression as well as bipolar disorder in low-resource users clearly underlining the research gap.



(a). Performance vs Knowledge augmentation

(b).Gain vs Knowledge augmentation

Fig. 1 Defining Low-Resource Users



(a). SVM

(b). LR

(c). CNN

(d). BERT

Fig. 2 Distribution of users on 20% data with respect to the established threshold line for respective methods.

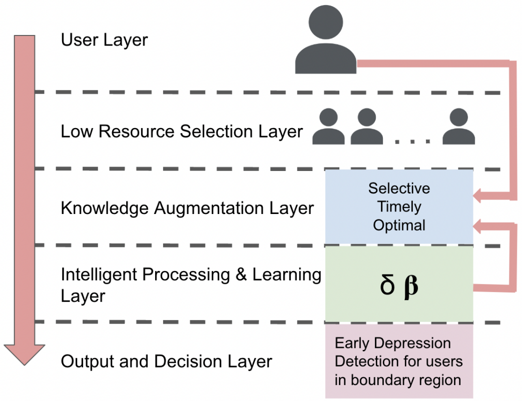

User Layer
Low Resource Selection Layer
Knowledge Augmentation Layer
Selective
Timely
Optimal
Intelligent Processing & Learning Layer
δ β
Output and Decision Layer
Early Depression Detection for users in boundary region

Fig. 3 System architecture

3 Research Questions

Identifying users in the boundary region remains a challenge particularly with early depression detection in low-resource users. Prior attempts assumed a vast number of individual posts, failed to focus on detecting boundary region users, and seldom capitalized on disorder specifics, e.g., bipolar disorder. Specifically, we address the following research questions:

RQ 1: Which users can be termed as low-resource users?
RQ 2: Is it possible to leverage knowledge augmentation to improve early depression detection in case of these low-resource users?
RQ 3: Can boundary region re-evaluation help in deciding type of depression?

4 Method: Soft Thresholding for Boundary Region Users (STBound)

At-risk, boundary-region users are currently not classified as depressed and in need of immediate attention, but rather, are kept under watch. Their immediate ongoing actions are potentially indicative of their inclinations on the depression spectrum. Considering limited future activity of these users can help determine their mental health status.

We propose an intelligent, selective and timely augmentation for the boundary region users. This approach uses the Soft Thresholding based Boundary detection method (STBound) to identify low-resource users on the brink of depression. We clearly define low-resource users by conducting through empirical study and mathematical modelling on RSDD dataset as depicted in figure 1.

Figure 1 (a) shows F1 score and (b) shows gain in F1 score per unit additional post for CNN. The red colored box in figure 1 (b) is the low-resource user zone marked from number of posts where rate of change in gain in F1 is highest to number of posts where rate of change in gain in F1 is lowest addressing RQ 1. From this behavior depicted in figure 1 which is common across all machine learning methods on RSDD dataset we infer the following. Users with 50 to 200 posts can be termed as 'low-resource users' where intelligent use of information yields better results.

The above inference is for RSDD dataset and bounds might vary with a different dataset. To simulate low-resource users with variable number of posts we create a distribution identical to the original data distribution but bounded in established low resource user bounds. These bounds are defined considering the region of high gain per unit additional posts as per Figure 1. For RSDD dataset we scale it by considering 20% of posts of each user and low-resource user bounds of 50 and 200. We decide the scaling percentage to be 20% as 20% of median number of posts lies perfectly between 50 and 200. As a result, we have successfully simulated a dynamic scenario of low-resource users with each user having number of posts between 50 and 200.

4.1 Hard Threshold Line

A hard threshold line is an empirically generated dynamic threshold separating users into depressed and not depressed categories. It is obtained by fitting linear regression to empirically determined threshold values on first 100, 200, 300 and 400 posts. The threshold values are obtained from validation set as per equation 1 where $F1[i]$ is the corresponding F1 score for threshold $th[i]$. Figure 2 depicts the distribution of users around respective hard threshold lines for methods SVM, LR, CNN and BERT, providing insights regarding users in the boundary region.

$$th_{ideal} = th[arg \max_i(F1[i])] \tag{1}$$

4.2 Soft Threshold Line

A soft threshold line, however is a variable line that determines the optimal region encompassing boundary users. It is obtained by subtracting β from the respective hard thresholds. Hard and soft threshold lines are specific to method under consideration. Hard threshold line helps us to get the threshold value dynamically – solely as a function of the number of posts a user has. All the users with score greater than their respective dynamic thresholds are classified as depressed users. In this case, the focus is on identifying the region where users are probably at risk but are not identified due to information inadequacy. Here soft threshold line plays the deciding role.

4.3 Boundary Region

Boundary region is defined as the region between hard and soft threshold lines. Users with scores lying in this boundary region are ‘at-risk boundary region users’. These users are re-evaluated with intelligent knowledge augmentation. Algorithm 1 gives the method for classifying a user based on selective re-evaluation and controlled increase in user’s posting data. It is carried out by increasing δ value step-wise until a termination criteria, i.e., δ_depth_count, is reached.

4.4 System Architecture

Figure 3 depicts an overview of the system architecture where user layer provides input to low resource user selection layer. The knowledge augmentation layer augments the input from previous layers and provide it to intelligent processing and learning layer. The output and decision layer performs early depression detection for boundary region users. We evaluated STBound with all four methods under consideration. STBound identifies low-resource users on the brink of depression successfully. Further, it improves early depression detection accuracy significantly with δ% controlled increase in user posts.

Algorithm 1 STBound algorithm

```
function EVALUATE($user, \delta, \beta, i$)
    $i \leftarrow i + 1$
    if $i == \delta\_depth\_count$ then
        RETURN(0)
                                                  ▷ User is not depressed
    else
        $posts \leftarrow$ SIM_LOW_RES($user, \delta$)
        $l \leftarrow$ LEN($posts$)
        $hard\_thr \leftarrow$ GET_THR($method, l$)
        $soft\_thr \leftarrow hard\_thr - \beta$
        $score \leftarrow$ METHOD($posts$)
        if $score > hard\_thr$ then
            RETURN(1)
                                                  ▷ User is depressed
        else if $score > soft\_thr$ then
            EVAL($user, \delta + inc, \beta, i$)
                                                  ▷ Boundary user
        else
            RETURN(0)
                                                  ▷ User is not depressed
        end if
    end if
end function
```

5 Analysis of Proposed Model

Let the initial number of posts of a low-resource user be p. We are adding δ% extra posts for re-evaluation. Define:

$$\gamma = 1 + \delta \tag{2}$$

Hence, the total number of posts for re-evaluation are:

$$p + p\delta = p\gamma \tag{3}$$

We model the score of the ML method as a function of number of posts of the low-resource user by:

$$y = 1 - \frac{1}{e^{\frac{x}{a}}} \tag{4}$$

where a is the depressed user behavior parameter. Here, the output score of the model is zero when the number of posts is zero and y $\longrightarrow$ 1 as the number of posts $\longrightarrow$ ∞. This curve is the most suited model for score as per empirical data. Intuitively, previously depicted sigmoid modeling for F1 score (figure 1 (a)) also makes sense as even though the score increases with increase in posts for every user, its reflection in F1 score will occur late, i.e., after significant data are obtained.

Define the hard threshold line by:

$$y = t^{‘}x + k \tag{5}$$

Proposition 1. *For a low-resource user who is depressed but is currently in the boundary region, lower bound on γ for extra posts needed for correct re-evaluation can be stated as:*

$$\gamma > \frac{1-k}{t^{‘}p} + \frac{W_{-1}\left(\frac{-e^{-\frac{(1-k)}{at^{‘}}}}{at^{‘}}\right)}{\frac{p}{a}} \tag{6}$$

where W_{-1} is the lower branch of Lambert function and p is the number of posts of the low-resource user.

Proof: For the low-resource user to be classified as depressed after δ% increase in posts, score by the model for $p(1+\delta) = p\gamma$ posts should be greater than corresponding threshold. Hence:

$$1 - \frac{1}{e^{\frac{p\gamma}{a}}} > t^{‘}p\gamma + k \tag{7}$$

Therefore,

$$\left(e^{\frac{-p}{a}}\right)^{\gamma} + (t^{‘}p)\gamma + (k-1) < 0 \tag{8}$$

Equation $a^x + bx + c = 0$ can be expressed as:

$$ln(a)\left(-x-\frac{c}{b}\right)e^{ln(a)\left(-x-\frac{c}{b}\right)} = ln(a)\frac{a^{\frac{-c}{b}}}{b} \tag{9}$$

This is of form $ze^z = k$ and can be solved using Lambert's W function: $z = W(k)$. Hence we have,

$$ln(a)\left(-x-\frac{c}{b}\right) = W\left(ln(a)\frac{a^{\frac{-c}{b}}}{b}\right) \tag{10}$$

Hence,

$$x = \frac{-c}{b} - \frac{W\left(ln(a)\frac{a^{\frac{-c}{b}}}{b}\right)}{ln(a)} \tag{11}$$

The roots of the equation on the left in (8) can be found by substituting $a = e^{\frac{-p}{a}}$, $b = t^{`}p$ and $c = k-1$ in (11):

$$root = \frac{1-k}{t^{`}p} + \frac{W\left(\frac{-e^{-\frac{(1-k)}{at^{`}}}}{at^{`}}\right)}{\frac{p}{a}} \tag{12}$$

Lower bound solution can be expressed using lower branch of Lambert's W function W_{-1} as:

$$\gamma > \frac{1-k}{t^{`}p} + \frac{W_{-1}\left(\frac{-e^{-\frac{(1-k)}{at^{`}}}}{at^{`}}\right)}{\frac{p}{a}} \tag{13}$$

$W_{-1}(.)$ can be evaluated using the Newton's method:

$$w_{j+1} = w_j - \frac{w_j e^{w_j} - z}{e^{w_j} + w_j e^{w_j}} \tag{14}$$

where w_0 for the lower branch can determined using Lajos Lóczi's formulation [26].

6 Experimentation

6.1 Datasets

The RSDD dataset [37] is a comprehensive dataset suitable for experimentation related to depression. It was created by annotating users from Reddit dataset[1] which is available publicly. RSDD is an extensive dataset spanning from Jan 2006 to Oct 2016.Using RSDD, we simulated low-resource platform users to facilitate experi-

[1] https://files.pushshift.io/reddit/

ments related to early depression detection. RSDD has 107,274 control and 9210 diagnosed users. Each user has 969 posts on an average with a mean length of 148 tokens. The dataset has three components: training, validation and testing. The SMHD dataset consists of Reddit posts of users who have claimed to have been diagnosed with one or several of nine mental health conditions ('diagnosed users'), and matched control users. It is a large dataset that covers diverse mental health conditions. It has a total number of 385,476 users consisting of 6434 bipolar, 14,139 depressed and 335,952 control users. On an average, control users have 310 posts, depressed users have 162 posts and bipolar users have 158 posts. We used this dataset to evaluate STBound performance to separate bipolar disorder by selecting depressed and bipolar users.

Table 1 STBound on low-resource RSDD users

	SVM				LR				CNN				BERT			
δ	β	F1	Pr	Rc	β	F1	Pr	Rc	β	F1	Pr	Rc	β	F1	Pr	Rc
0	0	0.321	0.327	0.311	0	0.388	0.419	0.361	0	0.446	0.431	0.462	0	0.422	0.425	0.419
0.1	0.0005	0.330	0.323	0.337	0.075	0.393	0.413	0.375	0.005	0.459	0.453	0.465	0.05	0.425	0.430	0.420
	0.0010	0.333	0.327	0.339	0.150	0.395	0.411	0.380	0.010	0.463	0.458	0.468	0.10	0.430	0.431	0.429
	0.0015	0.335	0.330	0.340	0.225	0.398	0.406	0.390	0.015	0.466	0.463	0.469	0.15	0.435	0.437	0.433
	0.0020	0.336	0.331	0.341	0.300	0.399	0.407	0.391	0.020	0.466	0.461	0.471	0.20	0.438	0.436	0.440
0.2	0.0005	0.332	0.326	0.338	0.075	0.398	0.418	0.380	0.005	0.478	0.482	0.474	0.05	0.427	0.429	0.425
	0.0010	0.342	0.336	0.348	0.150	0.403	0.414	0.393	0.010	0.480	0.483	0.477	0.10	0.435	0.443	0.427
	0.0015	0.344	0.331	0.358	0.225	0.406	0.407	0.405	0.015	0.482	0.484	0.480	0.15	0.442	0.451	0.433
	0.0020	0.346	0.328	0.366	0.300	0.408	0.404	0.412	0.020	0.482	0.484	0.480	0.20	0.452	0.460	0.444
0.3	0.0005	0.340	0.329	0.352	0.075	0.405	0.410	0.400	0.005	0.492	0.498	0.486	0.05	0.442	0.458	0.427
	0.0010	0.351	0.337	0.366	0.150	0.410	0.409	0.411	0.010	0.496	0.497	0.495	0.10	0.455	0.477	0.435
	0.0015	0.354	0.327	0.386	0.225	0.414	0.405	0.423	0.015	0.496	0.497	0.495	0.15	0.467	0.488	0.448
	0.0020	0.354	0.327	0.386	0.300	0.415	0.406	0.424	0.020	0.496	0.497	0.495	0.20	0.467	0.488	0.448
0.4	0.0005	0.343	0.334	0.352	0.075	0.408	0.406	0.410	0.005	0.506	0.514	0.498	0.05	0.448	0.468	0.430
	0.0010	0.352	0.331	0.376	0.150	0.415	0.410	0.420	0.010	**0.509**	0.517	0.501	0.10	0.463	0.487	0.441
	0.0015	0.357	0.322	0.400	0.225	0.420	0.410	0.431	0.015	0.509	0.517	0.501	0.15	**0.479**	0.509	0.452
	0.0020	**0.359**	0.315	0.418	0.300	**0.421**	0.411	0.432	0.020	0.509	0.517	0.501	0.20	0.479	0.509	0.452

6.2 Ethics and Privacy

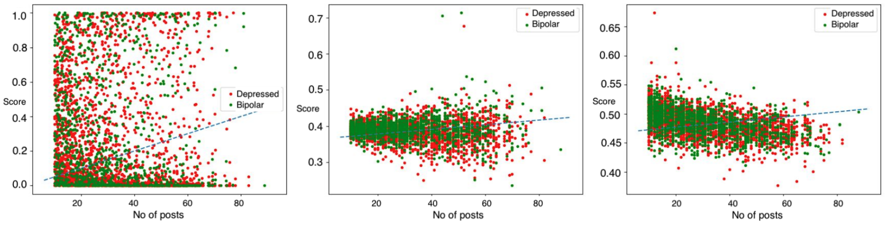
Personalized healthcare dataare sensitive. The data used are anonymized and due care is taken to minimize the risk while conducting experiments. The RSDD dataset contains the posts those are publicly available for academic use. All necessary care with reference to ethics and privacy was taken [37]. In this context, data were stored on secure servers, and no attempts were made to map, associate or re-identify users. Similar care was taken in the case of SMHD dataset also [7]. We refrain from making any details of these data publicly available. No attempt what so ever to link users to social media accounts was made.

6.3 Experimental Setup

We evaluate STBound on simulated low-resource users from RSDD dataset in combination with both traditional and connectionist machine learning methods. We considered lexical bag-of-words models Logistic Regression (LR) and SVM, conventional CNN and transformer based Bidirectional Encoder Representations from Transformers (BERT).

To assess the performance improvement of neural network based models over traditional machine learning models, we first evaluate the performance of STBound for LR and SVM. For these models, lexical bag-of-words (BoW) features were used as input. To minimize noise, we empirically determined a minimum document frequency of 12 for words to be considered. We fit the BoW tokenizer on the training set. For the CNN model, we had an embedding size of 50 per token. Additionally, the model had two Convolution1D layers with 25 filters, filter length of 3 and Rectified Linear Unit (ReLU) activation function. The model then had a 50 neuron dense layer and an output layer with Softmax activation function. A learning rate of 0.001 and 5 epochs led to the best results on the validation set.

RSDD has 969 posts per user on an average which results in memory constraints while running BERT. Considering these limitations, we considered 310x128 tokens per user and experimented with the following BERT [16] models: ‘small bert/bert_en_uncased_L-4_H-512_A-8’ and ‘small_bert/bert_en_uncased_L-2_H-128_A-2’. The first model is more intricate and hence we could accommodate only 175x128 tokens. Acknowledging this trade-off, we infer that the latter gives best results. Note that here L denotes the number of layers, i.e., transformer blocks. Also note that H denotes the hidden size while A denotes the number of self-attention heads. A learning rate of $2e-5$ and 3 epochs led to the best results on the validation set.



(a) LR (b) CNN (c) BERT

Fig. 4 STBound for separating Bipolar from Depressed users

6.4 Language Study

We conducted detailed language analysis to identify specific language characteristics which distinguish low-resource users from others. We observe that low-resource users have significantly higher usage of self-referencing phrases [4], i.e., 13% higher when calculated per unit sentence than others. Additionally we also notice significantly (i.e., 5%) lower occurrence of depression indicative phrases in low-resource users when compared with other users. Here we calculated depression indicative phrases [24] per unit sentence. Lower occurrence of depression indicative phrases in low-resource users justifies the need for thresholding and augmentation.

7 Results and Analysis

We have evaluated STBound on simulated low-resource users from RSDD dataset in combination with lexical bag-of-words, conventional CNN-based and transformer based ML methods. For comparison as a baseline, we obtain results on simulated low-resource users using dynamic thresholding. Here, users are clearly classified into depressed or not depressed using the obtained hard threshold. We use F1 score, Precision and Recall for comparing performance across different methods. We have focused on F1 score as Precision and Recall are equally important for depression detection. Table 1 shows the results of STBound with SVM, LR, CNN and BERT for different combinations of δ and β values. It adds δ% data to simulated low-resource users which satisfy the definition of boundary users. The boundary in this case is defined using the value of β. The variation in β values across models can be attributed to the differences in training algorithms and their respective parameters. Hence, optimal boundary region is determined separately for each method.

The highest F1 values are represented in bold in table 1. We note that in table 1, the first row with δ and β value equal to zero is our baseline and all performance improvements are represented with respect to it. We obtain an F1 of 0.359 for SVM for a δ value of 0.4 and β value of 0.002. Similarly we obtain an F1 of 0.421 for LR for δ value of 0.4 and β value of 0.3. In case of CNN, δ value of 0.4 and β value of 0.01 give the best F1 of 0.509. Further, in case of BERT, δ value of 0.4 and β value of 0.15 give the best F1 of 0.479. Here an important point to note is that BERT model was trained on limited data of each user due to memory constraints.

To compare CNN and BERT, we ran experiments in identical setups. Here, we consider 310x128 tokens per user for training and testing both models. From this we infer that BERT model outperforms CNN model in the identical setup. Experimental findings and analysis illustrate significant improvement in the F1 values with selective incorporation of boundary region re-evaluation with knowledge augmentation addressing RQ 2. An increase in δ value leads to an increase in F1. Similar trends are observed with an increase in β values, until a saturation level is reached. For example, for CNN, as β approaches 0.020, the results start saturating. Similar trends were observed across all the methods for the 18 β values evaluated. This indicates a relative separation between ‘at-risk boundary’ users and users who are not depressed. Additionally, for higher δ values, an increase in β value leads to greater improvements. With an optimal selection of δ and β values, we can get up to 8.51% improvement for SVM, up to 11.84% improvement for LR, up to 14.13% improvement for CNN and up to 13.51% improvement for BERT.

$$Efficacy = \frac{F1(\beta_{saturation}) - F1(\beta_0)}{\delta * F1(\beta_0)} \tag{15}$$

Table 2 Efficacy

δ	SVM	LR	CNN	BERT
0.1	**0.467**	**0.284**	**0.448**	**0.379**
0.2	0.389	0.258	0.404	0.356
0.3	0.343	0.232	0.374	0.356
0.4	0.296	0.213	0.353	0.338

The lower δ and β values with higher performance gain can help to identify the optimal boundary condition. We define efficacy as percentage improvement per unit δ. The obtained efficacy values can be found in table 2. For SVM and LR the highest efficacy values are 0.467 and 0.284 respectively. These values are observed at δ value 0.1. In case of CNN and BERT the highest efficacy values are 0.448 and 0.379 respectively and are also observed at δ value 0.1. In a real life setting, if we encounter boundary users with slow posting rate then optimal δ values can be determined using an efficacy index based on the urgency. Highest efficacy points can help us to fine-tune boundary region parameters.

8 Bipolar vs Depressed

We extended STBound to identify unipolar and bipolar depression. Misdiagnosis of bipolar depression is common. It is misdiagnosed as major depressive disorder (unipolar depression). Unipolar depression is treated by antidepressants while bipolar depression requires mood stabilizers [38]. It is important to identify bipolar depression because if missed, it will be treated like unipolar (with antidepressants) leading to increase in maniac episodes in patients – aggravating risks of self-harm and suicide multi-fold [19].

In early detection scenarios we have very limited data to detect bipolar depression. Higher percentage of bipolar cases lie in the boundary region defined by STBound. Re-evaluation of these users can help us to improve the classification of users in low-resource scenarios. Figure 4 depicts the distribution of users and use of STBound for separating bipolar and depressed users. Table 3 gives the performance of LR, CNN and BERT in separating bipolar disorder using hard threshold and using STBound. It is observed that performance of CNN gets a sharp leap using hard threshold. The performance further improves by 8.7% by use of STBound. In case of LR, after initial improvement obtained using hard threshold, STBound provides additional improvement of 18.6%. Similar trends are observed in case of BERT where after initial F1 score improvement, STBound provides 8.9% additional improvement. These results are indicative of possible use of STBound or it's enhancement to separate users with bipolar disorder addressing RQ 3. Though the improvements are realized at the cost of precision, the significant leap is definitely conclusive and promising for complex problem of separating users with bipolar depression.

Table 3 STBound on separating Bipolar from Depressed.
† Hard threshold, ‡ STBound, on low-resource users.

Method	F1	Precsion	Recall	Method	F1	Precsion	Recall	Method	F1	Precsion	Recall
LR	0.316	0.353	0.286	CNN	0.049	0.518	0.026	BERT	0.321	0.394	0.270
LR†	0.349	0.349	0.349	CNN†	0.482	0.480	0.485	BERT†	0.495	0.389	0.680
LR‡	**0.414**	0.333	0.547	CNN‡	**0.524**	0.378	0.854	BERT‡	**0.539**	0.381	0.921

9 Conclusion

We addressed RQ1 by defining low resource users empirically and mathematically. Depression detection for boundary region low-resource users is always an important and challenging task. The delay in depression detection for such users may result in delay in treatment and severe after effects. To address this issue, STBound performs selective intelligent knowledge augmentation and identifies boundary regions with higher precision. It further improves the accuracy of depression detection for the users in the boundary region by effective increase in δ value. The proposed method of selective and intelligent knowledge augmentation fetches improvement in overall F1 score on an average by 11.9% across all methods addressing RQ2. This substantial improvement helps in identifying the depressed boundary users on the brink of depression those otherwise would have gone unidentified. Early depression detection becomes even more crucial when it comes to separating bipolar disorder. Failure to separate bipolar disorder may result in delay in providing right treatment and further worsening the user's condition. STBound also improves the F1 score of separation of bipolar users by 12.1% addressing RQ3.

As a future work, other re-evaluation techniques can be combined with STBound. Fitting a curve with variable β values can optimally encompass the boundary region for further improvements. STBound with some contextual inputs can lead to promising techniques to separate bipolar disorder. Additionally, distribution of expressions over the time period can be used along with STBound to detect bipolar disorder in case of comorbidity.

References

1. Adrian, M., Coifman, J., Pullmann, M.D., Blossom, J.B., Chandler, C., Coppersmith, G., Thompson, P., Lyon, A.R.: Implementation determinants and outcomes of a technology-enabled service targeting suicide risk in high schools: Mixed methods study. JMIR Ment Health **7**(7), e16338 (2020). DOI 10.2196/16338. URL https://mental.jmir.org/2020/7/e16338
2. Agnihotri, N.: Review on machine learning techniques to predict bipolar disorder. TechRxiv (2021). URL https://doi.org/10.36227/techrxiv.14346050.v1
3. Bucur, A.M., Cosma, A., Dinu, L.P.: Early risk detection of pathological gambling, self-harm and depression using bert (2021)
4. Bucur, A.M., Podina, I., Dinu, L.: A psychologically informed part-of-speech analysis of depression in social media (2021). DOI 10.26615/978-954-452-072-4-024
5. Burdisso, S.G., Errecalde, M., Gómez, M.M.: A text classification framework for simple and effective early depression detection over social media streams. Expert Systems with Applications **133**, 182–197 (2019). DOI https://doi.org/10.1016/j.eswa.2019.05.023
6. CNN: More than 2,000 California mental health clinicians set to strike: CNN international. https://edition.cnn.com/2022/08/14/business/kaiser-mental-health-clinicians-strike/index.html (2022). Accessed 15-Aug-2022
7. Cohan, A., Desmet, B., Yates, A., Soldaini, L., MacAvaney, S., Goharian, N.: Smhd: A large-scale resource for exploring online language usage for multiple mental health conditions. In: Proceedings of the 27th International Conference on Computational Linguistics (COLING), p. 1485–1497. Association for Computational Linguistics (2018). URL https://www.aclweb.org/anthology/C18-1126
8. Cohan, A., Young, S., Goharian, N.: Triaging mental health forum posts. In: Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology, pp. 143–147. Association for Computational Linguistics, San Diego, CA, USA (2016). DOI 10.18653/v1/W16-0316. URL https://aclanthology.org/W16-0316
9. Cohan, A., Young, S., Yates, A., Goharian, N.: Triaging content severity in online mental health forums. Journal of the Association for Information Science and Technology (JASIST) (2018). DOI 10.1002/asi.23865. URL http://dx.doi.org/10.1002/asi.23865
10. Coppersmith, G., Fine, A., Crutchley, P., Carroll, J.: Individual differences in the movement-mood relationship in digital life data. In: Proceedings of the Seventh Workshop on Computational Linguistics and Clinical Psychology: Improving Access, pp. 25–31. Association for Computational Linguistics, Online (2021). DOI 10.18653/v1/2021.clpsych-1.3. URL https://aclanthology.org/2021.clpsych-1.3
11. Coppersmith, G., Hilland, C., Frieder, O., Leary, R.: Scalable mental health analysis in the clinical whitespace via natural language processing. In: 2017 IEEE EMBS International Conference on Biomedical Health Informatics (BHI), pp. 393–396 (2017). DOI 10.1109/BHI.2017.7897288
12. Coppersmith, G., Leary, R., Crutchley, P., Fine, A.: Natural language processing of social media as screening for suicide risk. Biomedical Informatics Insights **10**, 1178222618792860 (2018). DOI 10.1177/1178222618792860. URL https://doi.org/10.1177/1178222618792860. PMID: 30158822
13. Daly, M., Robinson, E.: Depression and anxiety during COVID-19. Lancet **399**(10324), 518 (2022)
14. Dattani, S., Ritchie, H., Roser, M.: Mental health. Our World in Data (2021). Https://ourworldindata.org/mental-health
15. Daveney, J., Panagioti, M., Waheed, W., Esmail, A.: Unrecognized bipolar disorder in patients with depression managed in primary care: A systematic review and meta-analysis. General hospital psychiatry **58**, 71–76 (2019)
16. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pp. 4171–4186. ACL, Minneapolis, Minnesota (2019). DOI 10.18653/v1/N19-1423. URL https://aclanthology.org/N19-1423

17. Elsafoury, F., Katsigiannis, S., Wilson, S.R., Ramzan, N.: Does BERT Pay Attention to Cyberbullying?, p. 1900–1904. Association for Computing Machinery, New York, NY, USA (2021). URL https://doi.org/10.1145/3404835.3463029
18. Erguzel, T.T., Sayar, G.H., Tarhan, N.: Artificial intelligence approach to classify unipolar and bipolar depressive disorders. Neural Computing and Applications **27**, 1607–1616 (2015)
19. Gitlin, M.J.: Antidepressants in bipolar depression: an enduring controversy. International Journal of Bipolar Disorders **6**(1), 25 (2018). DOI 10.1186/s40345-018-0133-9. URL https://doi.org/10.1186/s40345-018-0133-9
20. IHME: New global burden of disease analyses show depression and anxiety among the top causes of health loss worldwide, and a significant increase due to the covid-19 pandemic. IHME: Measuring what matters (2021). Accessed: 2022-01-05
21. Jan, Z., AI-Ansari, N., Mousa, O., Abd-alrazaq, A., Ahmed, A., Alam, T., Househ, M.: The role of machine learning in diagnosing bipolar disorder: Scoping review. J Med Internet Res **23**(11), e29749 (2021). DOI 10.2196/29749. URL https://www.jmir.org/2021/11/e29749
22. Kelly, K., Fine, A., Coppersmith, G.: Social media data as a lens onto care-seeking behavior among women veterans of the US armed forces. In: Proceedings of the Fourth Workshop on Natural Language Processing and Computational Social Science, pp. 184–192. Association for Computational Linguistics, Online (2020). DOI 10.18653/v1/2020.nlpcss-1.20. URL https://aclanthology.org/2020.nlpcss-1.20
23. Kulkarni, H., MacAvaney, S., Goharian, N., Frieder, O.: TBD3: A thresholding-based dynamic depression detection from social media for low-resource users. In: Proceedings of the Thirteenth Language Resources and Evaluation Conference, pp. 2157–2165. European Language Resources Association, Marseille, France (2022). URL https://aclanthology.org/2022.lrec-1.232
24. Kumar, A., Sharma, A., Arora, A.: Anxious depression prediction in real-time social data (2019). DOI 10.48550/ARXIV.1903.10222. URL https://arxiv.org/abs/1903.10222
25. Luján, M., Torres, A.M., Borja, A.L., Santos, J.L., Sotos, J.M.: High-precise bipolar disorder detection by using radial basis functions based neural network. Electronics **11**(3) (2022). DOI 10.3390/electronics11030343. URL https://www.mdpi.com/2079-9292/11/3/343
26. Lóczi, L.: Guaranteed- and high-precision evaluation of the lambert w function. Applied Mathematics and Computation **433**, 127406 (2022). DOI https://doi.org/10.1016/j.amc.2022.127406
27. MacAvaney, S., Desmet, B., Cohan, A., Soldaini, L., Yates, A., Zirikly, A., Goharian, N.: RSDD-time: Temporal annotation of self-reported mental health diagnoses. In: Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic, pp. 168–173. Association for Computational Linguistics, New Orleans, LA (2018). DOI 10.18653/v1/W18-0618. URL https://aclanthology.org/W18-0618
28. Pattisapu, N., Prabhu, N., Bhati, S., Varma, V.: Leveraging Social Media for Medical Text Simplification, p. 851–860. Association for Computing Machinery, New York, NY, USA (2020). URL https://doi.org/10.1145/3397271.3401105
29. Shi, L., Thiebaud, P., McCombs, J.S.: The impact of unrecognized bipolar disorders for patients treated for depression with antidepressants in the fee-for-services california medicaid (medi-cal) program. Journal of affective disorders **82**(3), 373–383 (2004)
30. Sidana, S., Mishra, S., Amer-Yahia, S., Clausel, M., Amini, M.R.: Health monitoring on social media over time. In: Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '16, p. 849–852. Association for Computing Machinery, New York, NY, USA (2016). DOI 10.1145/2911451.2914697. URL https://doi.org/10.1145/2911451.2914697
31. Soldaini, L., Goharian, N.: Quickumls: a fast, unsupervised approach for medical concept extraction. In: In MedIR Workshop SIGIR (2016)
32. Sotudeh, S., Goharian, N., Young, Z.: Mentsum: A resource for exploring summarization of mental health online posts. In: Proceedings of the Thirteenth International Conference on Language Resources and Evaluation (LREC). European Language Resources Association (ELRA) (2022)

33. Suen, P.J.C., Goerigk, S., Razza, L.B., Padberg, F., Passos, I.C., Brunoni, A.R.: Classification of unipolar and bipolar depression using machine learning techniques. Psychiatry Res. **27**, 1607–1616 (2021)
34. Vedula, N., Parthasarathy, S.: Emotional and linguistic cues of depression from social media. In: Proceedings of the 2017 International Conference on Digital Health, DH '17, p. 127–136. Association for Computing Machinery, New York, NY, USA (2017). DOI 10.1145/3079452.3079465. URL https://doi.org/10.1145/3079452.3079465
35. Villatoro-Tello, E., Ramírez-de-la Rosa, G., Gática-Pérez, D., Magimai.-Doss, M., Jiménez-Salazar, H.: Approximating the Mental Lexicon from Clinical Interviews as a Support Tool for Depression Detection, p. 557–566. Association for Computing Machinery, New York, NY, USA (2021). URL https://doi.org/10.1145/3462244.3479896
36. WHO: World health organization depression. World Health Organization News (2021). Accessed: 2021-30-12
37. Yates, A., Cohan, A., Goharian, N.: Depression and self-harm risk assessment in online forums. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp. 2968–2978. Association for Computational Linguistics, Copenhagen, Denmark (2017). DOI 10.18653/v1/D17-1322. URL https://aclanthology.org/D17-1322
38. Yatham, L.N., Kennedy, S.H., Parikh, S.V., Schaffer, A., Bond, D.J., Frey, B.N., Sharma, V., Goldstein, B.I., Rej, S., Beaulieu, S., Alda, M., MacQueen, G., Milev, R.V., Ravindran, A., O'Donovan, C., McIntosh, D., Lam, R.W., Vazquez, G., Kapczinski, F., McIntyre, R.S., Kozicky, J., Kanba, S., Lafer, B., Suppes, T., Calabrese, J.R., Vieta, E., Malhi, G., Post, R.M., Berk, M.: Canadian network for mood and anxiety treatments (CANMAT) and international society for bipolar disorders (ISBD) 2018 guidelines for the management of patients with bipolar disorder. Bipolar Disord **20**(2), 97–170 (2018)
39. Zafar, A., Chitnis, S.: Survey of depression detection using social networking sites via data mining. In: 2020 10th International Conference on Cloud Computing, Data Science Engineering (Confluence), pp. 88–93 (2020). DOI 10.1109/Confluence47617.2020.9058189