A Framework for Classifying Online Mental Health-Related Communities With an Interest in Depression

Budhaditya Saha, Thin Nguyen, Dinh Phung, and Svetha Venkatesh

***Abstract*—Mental illness has a deep impact on individuals, families, and by extension, society as a whole. Social networks allow individuals with mental disorders to communicate with others sufferers via online communities, providing an invaluable resource for studies on textual signs of psychological health problems. Mental disorders often occur in combinations, e.g., a patient with an anxiety disorder may also develop depression. *This co-occurring mental health condition provides the focus for our work on classifying online communities with an interest in depression*. For this, we have crawled a large body of 620 000 posts made by 80 000 users in 247 online communities. We have extracted the topics and psycholinguistic features expressed in the posts, using these as inputs to our model. Following a machine learning technique, we have formulated a joint modeling framework in order to classify mental health-related co-occurring online communities from these features. Finally, we performed empirical validation of the model on the crawled dataset where our model outperforms recent state-of-the-art baselines.**

***Index Terms*—Health Information Management, statistical learning, Predictive Models.**

I. Introduction

MENTAL health disorders are prevalent around the world. Estimates from the World Health Organization suggest that the lifetime prevalence of disorders in the Diagnostic and Statistical Manual of Mental Disorders is between 18% and 36% or between 10% and 19% of a population within a 12-month period [1]. An estimated 350 million people are affected by depression, and 800 000 people are thought to die by suicide each year [2]. The Internet offers a range of help-seeking options for individuals with a mental health disorder, including online communications which can offer peer support.

Recent studies have shown that machine learning and data mining techniques can be applied to online communities related to mental health [3]–[7]. Such techniques can analyze the content and linguistic styles of online discussions, and have been shown to differentiate communities interested in different mental health conditions [8]. A learning paradigm used in our previous work [7] considers each pair of outcomes to be modeled by a binary classifier. However, this is a relatively simplistic approach considering the relationship between topics and different communities. For example, "depression" may be relevant as a clinical condition in itself, may be present as a comorbidity with another disorder, or may reflect symptoms experienced as part of another diagnosable disorder, such as bipolar disorder. To reflect this, we formulate a joint learning framework where the model parameters of specific online communities are learned in an integrated manner. To achieve this, the cost function is composed of a suitable loss function and a correlation component that exploits the relatedness between these outcomes through an appropriate regularization. The framework jointly learns the relationship among related mental health communities interested in depression.

Manuscript received November 02, 2015; revised February 05, 2016 and March 02, 2016; accepted March 11, 2016. Date of publication March 18, 2016; date of current version July 06, 2016.

The authors are with the Centre for Pattern Recognition and Data Analytics, Faculty of Science and Technology, Deakin University, Vic. 3216, Australia (e-mail: budhaditya.saha@deakin.edu.au; thin.nguyen@deakin.edu.au; dinh.phung@deakin.edu.au; svetha.venkatesh@deakin.edu.au).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JBHI.2016.2543741

II. Background

A. Social Media as Sensors for Mental Health

Social media is quickly becoming a tool for health care analysis. Grajales *et al.* [9] has presented a guideline for clinicians, patients, and related stakeholders to use social media effectively while mitigating the risk factors. The online social media interventions have been used in the treatment of depression with a moderate success [10]. Park *et al.* [6] found that the depressive patients often use Twitter as a tool for social awareness and emotional interaction. The tweet activity is also used to predict 20 of 27 health-related statistics, including obesity and teen birth rates in USA [11]. Depression screening is conducted by Facebook, where subscribers' reveal symptoms of major depressive episodes [12], [13]. Blogging is found to improve maternal well being [14] by helping new mothers feel more connected to relatives and friends. Numerous online communities interested in health problems are found on the Live Journal blog site. Several studies on mood analysis using Live Journal data have been conducted in both personal and community settings [15], [16]. Live Journal communities interested in mental health concerns have also been investigated, for example, [5] for depression and [4] for autism.

B. Textual Features as Cues for Psychological Health

Recent studies have explored textual features as cues for mental health conditions [4], [5], [7]. For this type of cue, two main aspects have been popularly investigated: Topics and the language styles expressed in the text. Linguistic styles such as an expression of sadness or the use of swear words have been identified as the cues for depression [17]. Linguistic inquiry and word count (LIWC) [18] has been commonly used to capture language characteristics. LIWC features are considered to

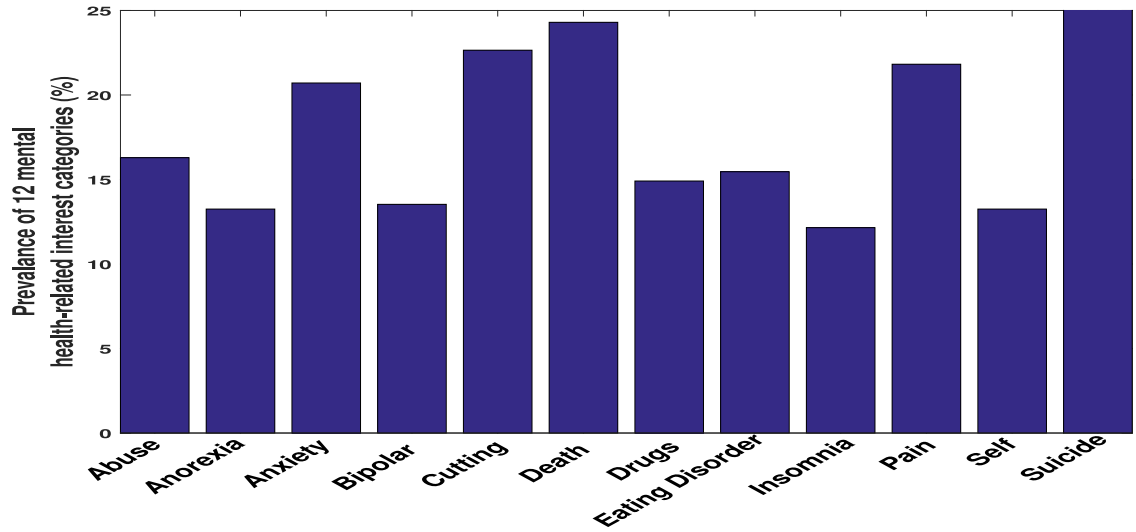

Fig. 1. Prevalence of 12 mental health-related interest categories in Live Journal communities

Topic	Word cloud	Topic	Word cloud
1	miss remember heart smile forget hold forever beautiful laugh kiss hug eyes	14	car house drive walk driving live town place
2	music favorite song live dance band listen rock songs listening kind color playing	15	change live world learn mind worth future living matter step realize past end moment happiness choice accept choose fight continue
3	away end leave far gone	16	eyes light sun dark hands sky rain blue black fire wind stars lips white beauty
4	livejournal site free pictures email check aim online myspace post add website html link picture computer interested	17	heart pain inside tears world mind broken words eyes soul death cry thoughts away dead head tear empty deep hold
5	community post support journal hello posting posted comment posts add comments join read reading joined	18	dear thank stop

Fig. 2. Indices and word clouds for selected topics.

be powerful predictors of mental health and depression-related disorders [19], [20]. Topics are extracted using popular Bayesian probabilistic modeling tools, such as latent Dirichlet allocation (LDA) [21]. LDA and its variants have been used previously to discover several mental ailments discussed in millions of tweets [22].

C. Joint Modeling of Multiple Problems

Recently, the joint modeling of multiple related problems has drawn significant interest. These methods exploit the commonality between related problems in order to learn efficient models. For example, multitask learning (MTL) is a joint learning method where an independent problem is considered to be a task and MTL computes parameters of multiple tasks in an integrated framework. Let us assume we have T supervised learning tasks and each task is associated with a linear predictive model f_t, where f_t is expressed as $f_t(\mathbf{x}) = \mathbf{w}_t^T \mathbf{x}_t$, where $\mathbf{w}_t$ and $\mathbf{x}_t$ are parameters and predictors of the task t. The MTL framework aims to enhance the predictive performance of a task by learning multiple related tasks simultaneously. Following this, the predictive functions f_t $\forall t{=}1,\ldots T$ are jointly learned by minimizing the following regularized empirical risk function:

$$\mathbf{w}_t^* = \min_{\mathbf{w}_t} \sum_{t=1}^{T} \frac{1}{n_t} \sum_{i=1}^{n_t} \mathcal{L}_i(\mathbf{w}_t, \mathbf{x}_t^i, y_t^i) + \lambda \mathcal{R}(\mathbf{w}_t), \qquad \forall t$$

where $\mathcal{L}_i$ is a loss function, $\mathcal{R}$ is a regularization function on $\mathbf{w}_t$ with regularization parameter λ, and n_t denotes the number of predictors in task t. Assuming tasks are related, MTL

techniques enforce several types of constraints on the task parameters $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_T]$ for modeling them jointly. For example, the multivariate distribution model [23] and the multilabel classification [24] model capture the shared structure among labels, while multioutcome regression [25] methods learn from the correlation between real-valued responses. In our paper, we model mental health-related comorbid communities in a MTL paradigm by considering each community as a task.

III. Methods

A. Dataset

We consider the online communities sourced from Live Journal where users can create and maintain a personal blog. People with similar interests can also form an "online community," a collective blog site in which multiple users can post and read messages on a common topic of interest. We initially identified communities with an interest in depression using the *"search communities by interest"* function on the website. The other interests of these communities were examined, and a subset of the most commonly occurring topics related to *"mental health with an interest in depression*" were selected. These included risk specific mental health disorders, risk factors, signs, symptoms, and outcomes. Inactive communities, those with fewer than 200 posts or no posts within the previous month, were excluded. This resulted in 247 communities with 12 major interest categories: Abuse, anorexia, anxiety, bipolar disorder, cutting, death, drugs, eating disorders, insomnia, pain, self-injury, and suicide (see Fig. 1).

B. Feature Extraction

Two feature sets were extracted during the experiments: Language style, using the LIWC package [18], and topics, by LDA [21]. The extracted features can be found here.[1]

1) The proportions of words in psycholinguistic categories as defined in the LIWC package were examined [18]. These categories were: Linguistic, social, affective, cognitive, perceptual, biological, relativity, personal concerns, and spoken.[2]
2) *Topics:* To extract topics, LDA [21] was used as a Bayesian probabilistic modeling framework. LDA extracts the probabilities $p\,(\text{vocabulary} \mid \text{topic})$—that is, words in a topic, and then assigns a topic to each word in a document. A Gibbs inference detailed in [26] was then implemented. The number of topics was set to 50 and the inference run for 5000 samples (see Fig. 2). The final Gibbs sample was then used to interpret the results.

C. A Framework for Classifying Communities Interested in Depression

In this section, we provide a framework for the joint modeling of communities interested in depression. Let us assume the feature matrix is denoted by $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]$ where D is the dimension of the feature space, i.e., $\mathbf{x}_i \in \mathcal{R}^D$. The matrix $\mathbf{Y} \in \mathcal{R}^{n \times m}$ denotes m outcomes of n data points. The outcomes are major "interest" categories where $m = 12$. The ith row of $\mathbf{Y}$ denoted by $\mathbf{Y}(i,:)$ represents all m outcomes of ith data point $\mathbf{x}_i$ whereas the kth column of $\mathbf{Y}$ denoted by $\mathbf{Y}(:,k)$ represents the kth outcome of n data points. For prediction problem,$\mathbf{Y}(i,k)$ takes a binary value from the set $\{1, -1\}$. Following a linear model, the relationship between the ith data point $\mathbf{x}_i$ and the corresponding kth outcome $\mathbf{Y}(i,k)$ is expressed as

$$\mathbf{Y}(i,k) = \mathbf{w}_k^T \mathbf{x}_i + b_k \qquad \forall i = 1, \ldots, n \tag{1}$$

where $\mathbf{w}_k$ and b_k are weight vector and bias component of the kth outcome, respectively. This model analogous to a *single-task learning* framework where the parameters of the kth outcome learns independently without considering correlations among the rest of the outcomes. In this framework, we capture the correlations among outcomes through the weight vectors in $\mathbf{W}$. Letting $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_m] \in \mathcal{R}^{D \times m}$ as a parameter matrix for m outcomes, we assume that the weight vectors are related through a prior distribution expressed as

$$p(\mathbf{W}) = \prod_{i=1}^{m} \mathcal{N}(\mathbf{w}_i | \mathbf{0}, \mathbf{I}_D) \mathcal{MN}_{D \times m}(\mathbf{W} | \mathbf{0}_{D \times m}, \mathbf{I}_D \otimes \mathbf{\Omega}) \tag{2}$$

where $\mathcal{N}(\mathbf{w}, \mathbf{\Lambda})$ denotes a multivariate normal distribution with mean $\mathbf{w}$ and covariance matrix $\mathcal{MN}_{D \times m}(\mathbf{A}, \mathbf{B} \otimes \mathbf{C})$ denotes a matrix-variate normal distribution with mean $\mathbf{A} \in \mathcal{R}^{D \times m}$, whereas $\mathbf{B} \in \mathcal{R}^{D \times D}$ is the row-covariance matrix, $\mathbf{C} \in \mathcal{R}^{m \times m}$ is column-covariance matrix, and $\otimes$ denotes the Kronecker product. The first component in (2) $(\mathcal{N}(\mathbf{w}_i | \mathbf{0}, \mathbf{I}_D))$ models the weight vector $\mathbf{w}_i$ individually and the second component $\mathcal{MN}_{D \times m}(\mathbf{W} | \mathbf{0}_{D \times m}, \mathbf{I}_D \otimes \mathbf{\Omega})$ learns the relatedness among the weight vectors through the covariance matrix $\mathbf{\Omega}$. The likelihood of the model for n number of i.i.d observations is

$$\prod_{i=1}^{n} p(\mathbf{y}_i | \mathbf{x}_i, \mathbf{W}, \mathbf{b}) = \prod_{i=1}^{n} \mathcal{N}(\mathbf{y}_i | \mathbf{W}^T \mathbf{x}_i + \mathbf{b}, \epsilon_i^2 \mathbf{I}_m) \tag{3}$$

where the posterior distribution of $\mathbf{W}$ is given by

$$p(\mathbf{W} | \mathbf{X}, \mathbf{Y}, \mathbf{b}, \mathbf{\Omega}, \boldsymbol{\epsilon}) \propto \prod_{j=1}^{n} p(\mathbf{y}_i | \mathbf{x}_i, \mathbf{W}, \mathbf{b}) p(\mathbf{W}). \tag{4}$$

The coefficient vectors $\mathbf{W}$, and parameters $\mathbf{b}$ and $\mathbf{\Omega}$ are obtained by minimizing the *maximum a posteriori* formulation, expressed as:

$$\min_{\mathbf{W}, \mathbf{b}, \mathbf{\Omega}} \sum_{i=1}^{n} \left[\sum_{k=1}^{m} \mathcal{L}_i(\mathbf{w}_k, \mathbf{x}_i, y_i^k, b_k) + \frac{\lambda_1}{2} \sum_k \mathbf{w}_k^T \mathbf{w}_k \right.$$
$$\left. + \frac{\lambda_2}{2} \mathrm{tr}(\mathbf{W} \mathbf{\Omega}^{-1} \mathbf{W}^T) + D \log |\mathbf{\Omega}| \right] \tag{5}$$
$$\text{s.t.}\, \mathbf{\Omega} \succeq \mathbf{0} \tag{6}$$

where $\mathcal{L}_i(\mathbf{w}_k, \mathbf{x}_i, y_i^k, b_k) = \frac{1}{2} \left(y_i^k - \mathbf{w}_k^T \mathbf{x}_i - b_k \right)^2$, λ_1 and λ_2 are regularization parameters. In the above formulation, the term

[1] http://bit.ly/1KS4CMr
[2] http://www.liwc.net/descriptiontable1.php, retrieved January 2015.

TABLE I
CLASSIFICATION PERFORMANCE OF 12 "INTEREST" CATEGORIES FROM TOPIC FEATURES

Mental Health Related Interest Categories	Single Task Learning	MTL	Proposed Framework
Abuse	0.818 (0.002)	**0.820**(0.003)	**0.820**(0.002)
Anorexia	0.882 (0.008)	0.884 (0.008)	**0.901**(0.010)
Anxiety	0.827 (0.009)	**0.829**(0.006)	**0.829**(0.008)
Bipolar	0.874 (0.007)	0.878 (0.005)	**0.887**(0.005)
Cutting	0.863 (0.004)	**0.864**(0.008)	0.863 (0.005)
Death	0.909 (0.005)	0.921 (0.006)	**0.930**(0.003)
Drugs	0.820 (0.007)	0.823 (0.006)	**0.832**(0.005)
Eating disorders	0.905 (0.008)	0.910 (0.007)	**0.939**(0.010)
Insomnia	0.816 (0.009)	0.828 (0.009)	**0.879**(0.005)
Pain	**0.825**(0.007)	0.824 (0.006)	0.824 (0.006)
Self-injury	**0.863**(0.008)	**0.863**(0.004)	0.862 (0.007)
Suicide	**0.840**(0.006)	0.839 (0.005)	0.838 (0.007)
Average	0.853 (0.007)	0.856 (0.006)	**0.870**(0.006)

The evaluation is done in terms of mean AUC. Numbers in bracket denotes standard deviation. The highest performances are highlighted. Numbers in bracket denotes standard deviation.

TABLE II
CLASSIFICATION PERFORMANCE OF 12 "INTEREST" CATEGORIES FROM TOPIC FEATURES

Mental Health-Related Interest Categories	Single Task Learning	MTL	Proposed Framework
Abuse	0.761 (0.009)	0.760 (0.010)	0.786(0.006)
Anorexia	0.839 (0.010)	0.838 (0.004)	0.840(0.008)
Anxiety	0.763 (0.011)	0.763 (0.009)	0.785(0.008)
Bipolar	0.789 (0.008)	0.789 (0.008)	0.817(0.012)
Cutting	0.818 (0.005)	0.818 (0.010)	0.823(0.010)
Death	0.867 (0.010)	0.867 (0.008)	0.871(0.009)
Drugs	0.758 (0.009)	0.770 (0.012)	0.793(0.012)
Eating disorders	0.858 (0.008)	0.861 (0.010)	0.868(0.006)
Insomnia	0.755 (0.008)	0.760 (0.011)	0.781(0.013)
Pain	0.768 (0.005)	0.766 (0.007)	0.782(0.010)
Self-injury	0.821 (0.006)	0.822 (0.005)	0.823(0.005)
Suicide	0.800 (0.010)	0.800 (0.010)	0.805(0.010)
Average	0.799 (0.008)	0.801 (0.008)	0.815(0.011)

from LIWC features. The evaluation is done in terms of mean AUC. Numbers in bracket denotes standard deviation. The highest performances are highlighted.

TABLE III
CLASSIFICATION PERFORMANCE OF 12 "INTEREST" CATEGORIES FROM TOPIC FEATURES

Mental Health-Related Interest Categories	Single Task Learning	MTL	Proposed Framework
Abuse	0.838 (0.008)	0.830 (0.007)	0.848(0.005)
Anorexia	0.898 (0.007)	0.891 (0.007)	0.898(0.007)
Anxiety	0.834 (0.006)	0.833 (0.007)	0.843(0.008)
Bipolar	**0.899**(0.008)	0.891 (0.007)	0.899(0.008)
Cutting	0.886 (0.008)	0.886 (0.007)	0.899(0.008)
Death	0.888 (0.010)	**0.921**(0.007)	0.921(0.006)
Drugs	**0.846**(0.006)	**0.846**(0.007)	0.846(0.005)
Eating disorders	0.921 (0.009)	0.918 (0.007)	0.932(0.010)
Insomnia	0.825 (0.010)	0.825 (0.007)	0.845(0.009)
Pain	0.832 (0.008)	**0.842**(0.007)	0.842(0.006)
Self-injury	0.877 (0.009)	0.871 (0.007)	0.887(0.008)
Suicide	0.844 (0.005)	0.848 (0.007)	0.854(0.005)
Average	0.864 (0.008)	0.866 (0.007)	0.876(0.008)

from augmented LIWC and topics features: The evaluation is done in terms of mean AUC. The highest performances are highlighted. Numbers in bracket denotes standard deviation.

$\text{tr}(\mathbf{W}\mathbf{\Omega}^{-1}\mathbf{W}^T)$ act as a coupling pairs among the weight vectors and also control the amount of sharing between any pair of the weight vectors. The positive definite constraints on $\mathbf{\Omega}$ in (6) is imposed to make it nonsingular. By assuming that both $\mathbf{W}$ and $\mathbf{\Omega}$ are sparse, we add a ℓ_1- norm regularization over both $\mathbf{W}$ and $\mathbf{\Omega}$, respectively. We can rewrite the formulation in (5) as

$$\min_{\mathbf{W},\mathbf{b},\mathbf{\Omega}^{-1}} \sum_{i=1}^{n}\Bigg[\sum_{k=1}^{m}\mathcal{L}_i(\mathbf{w}_k,\mathbf{x}_i,y_i^k,b_k)+\frac{\lambda_1}{2}\text{tr}(\mathbf{W}\mathbf{W}^T) + \frac{\lambda_2}{2}\text{tr}(\mathbf{W}\mathbf{\Omega}^{-1}\mathbf{W}^T)-D\log|\mathbf{\Omega}^{-1}| + \lambda_3||\mathbf{\Omega}^{-1}||_1+\lambda_4||\mathbf{W}||_1\Bigg]. \quad (7)$$

The inverse of the covariance matrix $\mathbf{\Omega}$ naturally satisfy the positive definite constraints on $\mathbf{\Omega}$. The formulation in (7) is not jointly convex with respect to all variables, however, it is convex with respect to each individual variable when the remaining variables are fixed. We use an efficient block coordinate descent method [27] to find optimal solutions of $\mathbf{W}$, $\mathbf{b}$ and $\mathbf{\Omega}^{-1}$, respectively.

IV. RESULTS

In this section, we evaluate the performance of the proposed model. The Live Journal data contains 620 060 posts from 78 647 users. We extracted LIWC and topic features from the online posts, which are 68- and 50-dimensional, respectively. We represent each post by an input feature vector (LIWC or Topics) and an outcome vector (interest categories) of dimension 12. The element the outcome vector is 1 or -1 depending on the presence or absence of an interest category. We have performed three experiments on the extracted data as follows:

1) First, we train a model using the topic features as an input and the 12 interest categories as outcomes.
2) Second, we use the LIWC features as an input and the 12 interest categories as outputs.
3) Finally, we combine the topics and LIWC features into a single feature set and repeat the experiment as before.

The efficacy of this proposed model was then evaluated against single-task logistic (STL) regression [28] and a MTL [29] framework. The results of these experiments and comparisons are summarized in Tables I–III.

A. Classification of Mental Health-Related "Interest" Categories Using Topic Features

We randomly partitioned the topic features into training, validation, and test sets in the ratio 70:20:10. Using the training set, we trained our model by pairing the input with expected output. The validation set was used in order to estimate how well the model has been trained and to estimate the model parameters and properties. We then applied the newly developed model to the test data and recorded the results. We repeated this experiment on 20 randomly partitioned datasets and present the average results in Table I, which shows the

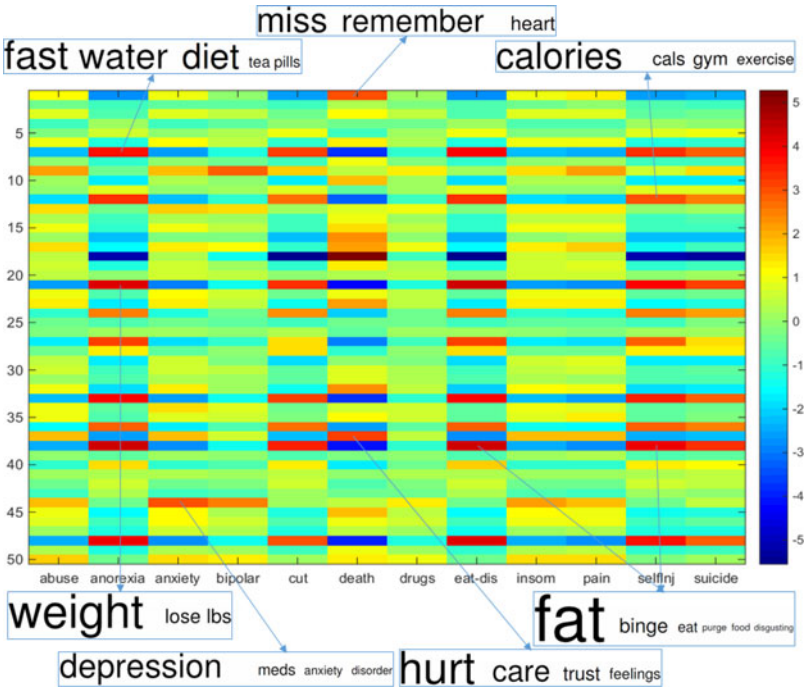

Fig. 3. Parameter matrix **W** estimated from the topic features: Each column of **W** denotes a weight vector of an "interest" category. Representative topic features are mentioned.

TABLE IV
SENSITIVITY AND SPECIFICITY ANALYSIS ON AUGMENTED LIWC AND TOPIC DATASET

	Single Task Learning		MTL		Proposed Framework	
Mental Health-Related Interest Categories	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
Abuse	72.8	81.6	83.1	84.1	**87.2**	**88.2**
Anorexia	74.5	82.2	84.7	86.7	**86.9**	**90.1**
Anxiety	78.9	86.2	83.3	89.8	**86.1**	**92.3**
Bipolar	85.9	**87.4**	85.9	**87.4**	**86.0**	87.1
Cutting	77.7	87.5	83.3	92.4	**84.4**	**95.1**
Death	75.1	81.3	81.6	**89.1**	82.3	**89.1**
Drugs	79.9	88.3	79.7	**90.6**	**80.0**	**90.6**
Eating disorders	73.3	83.5	78.1	91.7	**81.5**	**93.4**
Insomnia	75.9	80.4	80.2	87.1	**81.1**	**88.4**
Pain	80.1	87.7	**84.5**	95.9	**84.5**	**96.6**
Self-injury	79.6	86.1	81.8	90.2	**83.1**	**92.5**
Suicide	81.1	84.4	85.5	89.0	**87.2**	**93.8**
Average	77.9	84.7	82.6	89.5	**84.2**	**91.4**

The highest performances are highlighted.

performance of the proposed model against state-of-the-art baseline methods.

The proposed model outperforms the STL regression model for *9 out of 12* "interest" categories. In case of MTL, the proposed framework performs better for*8 out of 12* categories. In particular, for "Drugs," the proposed model (AUC 0.832) improves by a margin of 2% with respect to the STL model (AUC 0.820) and 1.8% with respect to the MTL model (AUC 0.823); for "Insomnia," the proposed model (AUC 0.839) improves by a margin of 3% with respect to the STL model (AUC 0.816) and 2% with respect to the MTL model (AUC 0.828). For all outcomes, the mean AUC of the proposed model (0.870) improved by a margin of $\approx 2\%$ and 1.9% with respect to the single-task and multitask model, respectively.

Notably for "Pain," "Self-Injury," and "Suicide," the performances are more or less similar to the single-task model (STL). We have examined the correlations of these categories with remaining others and we have found that these categories have low correlations (0.09–0.15). This finding is consistent with the properties of single-task models, which generally perform better with low-correlated outputs.

B. Classification of Mental Health-Related 'Interest" Categories Using LIWC Features

Table II shows the results from the LIWC feature sets. For "Drugs," the proposed model (AUC 0.793) improves by a margin of 4.5% with respect to the STL (AUC 0.758) and 3% with

TABLE V
ESTIMATED PREDICTIVE FEATURES FOR 12 MENTAL HEALTH-RELATED CATEGORIES INTERESTED IN DEPRESSION

Mental Health-Related Interest Categories	Topic Features	LIWC features
Abuse	hurt care trust feelings, cutting blood scars, bitch hell act wrong shut funny bullshit	you, negation, negative emotion, wc (word count)
Anorexia	fast water diet, calories gym exercise, weight lose lbs	ingest
Anxiety	depression anxiety medication, self suicide emotional, hurt care trust feelings	positive emotion. negation, word count (wc)
Bipolar	cutting blood scars, depression meds anxiety disorder, heart pain inside	sad, death
Cutting	cutting blood scars, fat binge eat purge food disgusting, eat food eating dinner	ingest, feel
Death	miss remember heart, hurt care trust feelings, mom dad family mother parents sister brother died	social, you, we, affect, positive emotion
Drugs	depression anxiety disorder medication, heart pain inside tears, suicide emotional thoughts	death, sad , negative emotion
Eating disorder	eat food dinner lunch, calories low soup dinner salad apple milk, eating recovery disorder healthy food.	ingest, biological process (bio), work
Insomnia	depression meds anxiety disorder, hurt care trust feelings	word count (wc), you, she-he
Pain	cutting blood arm deep pain razor wrist hurt suicide knife blade, heart pain inside tears, depression anxiety disorder	death, you, world count (wc)
Self-injury	calories gym exercise hour burn run, eating recovery disorder healthy anorexia,	injest, feel, biological process (bio)
Suicide	stay strong girls luck ladies, mum christmas lots loads ages, seriously hell crap sucks freaking horrible mood	injest, feel, biological process (bio)

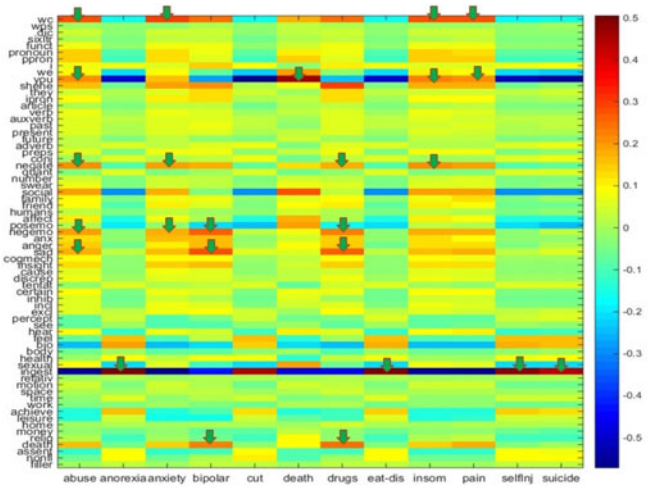

Fig. 4. Parameter matrix $\mathbf{W}$ estimated from the LIWC features: The columns of $\mathbf{W}$ represent weight vectors of "interest" categories. Down arrows indicate the representative features for an "interest." For example, "negative emotion" "negemo" is strongly found in "abuse," "anxiety," "bipolar," and "drugs" categories, respectively.

respect to MTL (AUC 0.770); while in the case of "Insomnia," the proposed model (AUC 0.781) improves by a margin of $\approx 3\%$ with respect to the STL (AUC 0.755) and 2.76% with respect to MTL (AUC 0.760). The mean AUC of the proposed model is improved by a margin of 2% and 1.75% with respect to single-task and multitask model. Our proposed model outperforms both STL and MTL models for all 12 "interest" categories.

C. Classification of Mental Health-Related 'Interest" Categories Using Augmented Topic and LIWC Features

In this experiment, we augment both LIWC and Topic features to a single dataset. The combined feature dimension is 118. Table III presents the performance of the proposed model on the combined data. The proposed model is better for *10 out of 12* interests with respect to single-task (STL) and multitask (MTL) model respectively. In particular, in case of "Abuse," the proposed model (AUC 0.848) improves by a margin of 2.17% with respect to the single-task model (AUC 0.838), while in case of "Insomnia," the proposed model (AUC 0.845) improves by a margin of 2.5% with respect to the single-task model (AUC 0.825). However, for "Bipolar" and "Drugs," the performance of the proposed model is same with single task model.

Fig. 3 shows the parameter matrix $\mathbf{W}$ of topic features where each column of $\mathbf{W}$ denotes a weight vector related to an outcome. For each interest category, the top three features are shown in Table V. Notably, some features are common across the categories. For example, "cut," "blood," and "scars" are important topics in "Abuse," "Bipolar," "Cutting," and "Pain" categories; "hurt," "care," "trust," and "feelings" are mostly discussed in "Abuse" and "Death" categories; "ingest" (i.e., taking food into the body) is a strong feature in "Anorexia," "Eating Disorder," and "Suicide.".

Fig. 4 shows the parameter matrix $\mathbf{W}$ for LIWC features. The representative features across interest categories are indicated by the down arrows. For example, "negemo" (negative emotion) is strongly present in "Abuse," "Anxiety," "Bipolar" and "Drugs," respectively; "ingest" is found in "Anorexia," "Cut," "Eating Disorder," "Self-Injury," and "Suicide."

Table VI shows a comparative performance of single-task (STL) and the proposed framework on LIWC, Topics, and the augmented dataset, respectively. The performance of the single-task model (STL) significantly improves on augmented dataset (mean AUC 0.864) with respect to LIWC (AUC 0.799) and topic features (AUC 0.853). We also observed that the mean AUC of the proposed model for the augmented data (0.870) outperforms the performance of LIWC features (AUC 0.799)

TABLE VI
COMPARATIVE PERFORMANCE OF 12 "INTEREST" CATEGORIES ON LIWC, TOPICS, AND AUGMENTED (LIWC + TOPICS) DATASETS, RESPECTIVELY

Mental Health-Related Interest Categories	Single-Task Learning (LIWC)	Single-Task Learning (Topics)	Single-Task Learning (LIWC + Topics)	Proposed (LIWC)	Proposed (Topics)	Proposed Framework (LIWC + Topics)
Abuse	0.761 (0.011)	0.818 (0.008)	0.838 (0.008)	0.786 (0.010)	0.820 (0.011)	**0.848**(0.008)
Anorexia	0.839 (0.008)	0.882 (0.007)	0.898 (0.006)	0.840 (0.006)	**0.901**(0.009)	0.898 (0.010)
Anxiety	0.763 (0.006)	0.827 (0.006)	0.834 (0.007)	0.785 (0.006)	0.829 (0.006)	**0.843**(0.008)
Bipolar	0.789 (0.005)	0.874 (0.006)	0.899 (0.007)	0.817 (0.007)	0.887 (0.007)	**0.899**(0.006)
Cutting	0.818 (0.011)	0.863 (0.012)	0.886 (0.012)	0.823 (0.010)	0.863 (0.010)	**0.899**(0.011)
Death	0.867 (0.012)	0.909 (0.011)	0.888 (0.015)	0.871 (0.015)	**0.930**(0.012)	0.921 (0.014)
Drugs	0.758 (0.013)	0.820 (0.008)	0.846 (0.009)	0.793 (0.008)	0.832 (0.009)	**0.846**(0.009)
Eating disorders	0.858 (0.016)	0.905 (0.010)	0.921 (0.012)	0.868 (0.012)	**0.939**(0.011)	0.931 (0.010)
Insomnia	0.755 (0.013)	0.816 (0.008)	0.825 (0.013)	0.781 (0.014)	**0.879**(0.011)	0.845 (0.013)
Pain	0.768 (0.011)	0.825 (0.009)	0.832 (0.010)	0.782 (0.014)	0.824 (0.015)	**0.842**(0.016)
Self-injury	0.821 (0.021)	0.863 (0.020)	0.877 (0.018)	0.823 (0.017)	0.862 (0.011)	**0.887**(0.009)
Suicide	0.800 (0.015)	0.840 (0.012)	0.844 (0.017)	0.805 (0.017)	0.838 (0.018)	**0.854**(0.016)
Mean AUC	0.799 (0.009)	0.853 (0.015)	0.864 (0.016)	0.815 (0.015)	0.870 (0.015)	**0.876**(0.014)

The evaluation is done in terms of mean AUC. The highest performances are highlighted. Numbers in bracket denotes standard deviation.

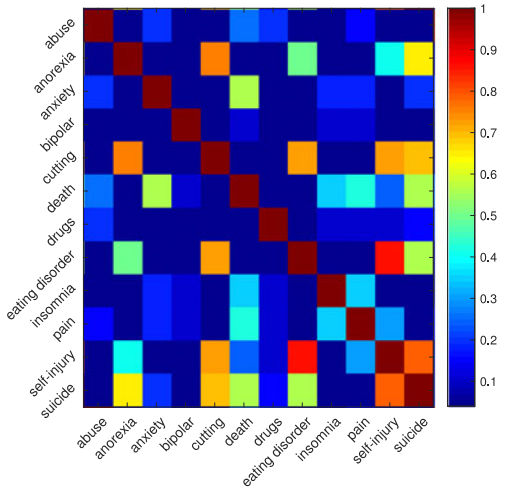

Fig. 5. Correlation matrix estimated from the augmented (LIWC and Topics) dataset for 12 "interest" categories.

and topic features (AUC 0.853), respectively. Hence, we have demonstrated that the predictive power of the augmented dataset is more than the LIWC and Topic features.

Fig. 5 shows the mean correlation matrix of 12 outcomes. In the case of the "Bipolar" and "Drugs," the correlations are low (0–0.15). Hence, the performance of the proposed model is similar to a single-task (STL) model. Overall, the multitask model (MTL) with the mean AUC 0.866 is better than the single-task model (STL) (mean AUC 0.864), however, the proposed model (mean AUC 0.876) outperforms both of them. Table IV shows the performance of the models in terms of sensitivity and specificity. As before, the proposed model outperforms the baselines.

V. DISCUSSION

This paper aimed at classifying *mental health-related communities with an interest in depression*. For this, we have found communities with an interest in depression using the "search communities by interest" function on the Live Journal website. The several interests of online communities were examined and a subset of the most commonly occurring topics related to depression was selected. The 12 identified categories are "Abuse," "Anorexia," "Anxiety," "Bipolar," "Cutting," "Death," "Drugs," "Eating Disorders," "Insomnia," "Pain," "Self-injury," and "Suicide."

Following machine learning and statistical methods, we focus on the two important aspects of mental health communities: The content of the topics posted, and the psycholinguistic processes used in these topics. We have extracted the topics discussed in online posts using Bayesian LDA and then selected the top most 50 topics discussed in the communities. The psycholinguistic processes are extracted using the LIWC package. We have formulated our linear regression model considering the topics and linguistic styles as inputs while the "interest" categories are outcomes. The entries of the outcome vector are 1 or -1 depending on the presence or absence of an interest category. We follow a joint learning approach where the parameters of the outcomes are learned in an integrated framework. We have performed the empirical validation of the model on the crawled dataset and evaluated the performance of the proposed model against recent state-of-the-art baselines.

Results indicate that the distinct topics and linguistic styles have a strong predictive power to classify mental health-related communities with an interest in depression. The latent topics are found to have a greater predictive power than linguistic features. This can be seen by comparing the third column in Tables II and III, respectively. Moreover, the performance of the proposed joint learning model outperforms state-of-the-art single-task (STL) and multitask (MTL) learning baselines. In case of topic features, the mean AUC of the proposed model is improved by a margin of 2% and 1.9%, whereas, in the case of linguistic features, the mean AUC is improved by 2% and 1.75% with respect to STL and MTL, respectively. The significant finding of this paper is that by augmenting both topics and linguistic features into a single feature representation, the classification performance improves significantly over state-of-the-art baselines.

This finding indicates that the topics discussed in a community are also relevant to other groups. This result confirms

that the discussion topics in online communities is extended beyond the "depressed feelings" and can be more specifically related to a range of other "interests" like "bipolar," "eating disorders," "emotional experience" (death, loss of a loved one), or "behaviours" (cutting, self-harm). The predictive features from linguistic analysis are consistent with topic analysis, for example, "Abuse" contains topic features such as "hurt," "care," "trust," and "feelings," "scars," and LIWC features like "you," "negation," "negative emotion" (see Table V).

VI. Conclusion

This study demonstrates that the linguistic features and topics discussed among the online communities have the potential to capture the mental status and presence of mental health-related communities. A number of significant examples were found where these features have strong indicative powers in the prediction of co-occurring communities interested in depression. This result shows the potential of social media and online communities in the early screening and monitoring of mental health-related communities with an interest in depression.

References

[1] R. C. Kessler *et al.*, "The global burden of mental disorders: An update from the WHO World Mental Health (WMH) surveys," *Epidemiologia e Psichiatria Sociale*, vol. 18, no. 01, pp. 23–33, 2009.

[2] World Health Organization, "Depression [fact sheet no. 369]," 2013.

[3] M. D. Choudhury, M. Gamon, S. Counts, and E. Horvitz, "Predicting depression via social media," in *Proc. Int. AAAI Conf. Weblogs Soc. Media*, 2013, pp. 128–137.

[4] T. Nguyen, T. Duong, S. Venkatesh, and D. Phung, "Autism blogs: Expressed emotion, language styles and concerns in personal and community settings," *IEEE Trans. Affective Comput.*, vol. 6, no. 3, pp. 312–323, Jul./Sep. 2015.

[5] T. Nguyen, D. Phung, B. Dao, S. Venkatesh, and M. Berk, "Affective and content analysis of online depression communities," *IEEE Trans. Affective Comput.*, vol. 5, no. 3, pp. 1949–3045, Jul./Sep. 2014.

[6] M. Park, D. McDonald, and M. Cha, "Perception differences between the depressed and non-depressed users in Twitter," in *Proc. AAAI Int. Conf. Weblogs Soc. Media*, 2013, pp. 476–485.

[7] T. Nguyen, B. O'Dea, M. Larsen, D. Phung, S. Venkatesh, and H. Christensen, "Differentiating sub-groups of online depression-related communities using textual cues," in *Proc. Int. Conf. Web Inform. Syst. Eng.*, 2015, pp. 216–224.

[8] T. Nguyen, B. O'Dea, M. Larsen, D. Phung, S. Venkatesh, and H. Christensen, "Using linguistic and topic analysis to classify sub-groups of online depression communities," *Multimedia Tools Appl.*, pp. 1–24, vol. 00, 2015.

[9] F. J. Grajales III, S. Sheps, K. Ho, H. Novak-Lauscher, and G. Eysenbach, "Social media: A review and tutorial of applications in medicine and health care," *J. Med. Internet Res.*, vol. 16, no. 2, 2014, Art. no. e13.

[10] S. M. Rice *et al.*, "Online and social networking interventions for the treatment of depression in young people: A systematic review," *J. Med. Internet Res.*, vol. 16, no. 9, 2014, Art. no. e206.

[11] A. Culotta, "Estimating county health statistics with Twitter," in *Proc. Annu. ACM Conf. Human Factors Comput. Syst.*, 2014, pp. 1335–1344.

[12] M. A. Moreno *et al.*, "Feeling bad on Facebook: Depression disclosures by college students on a social networking site," *Depression Anxiety*, vol. 28, no. 6, pp. 447–455, 2011.

[13] Y. Soo Jeong *et al.*, "Using online social media, Facebook, in screening for major depressive disorder among college students," *Int. J. Clin. Health Psychol.*, vol. 13, no. 1, pp. 74–80, 2013.

[14] B. T. McDaniel, S. M. Coyne, and E. K. Holmes, "New mothers and media use: Associations between blogging, social networking, and maternal well-being," *Maternal Child Health J.*, vol. 16, no. 7, pp. 1509–1517, 2012.

[15] T. Nguyen, D. Phung, B. Adams, and S. Venkatesh, "Mood sensing from social media texts and its applications," *Knowl. Inform. Syst.*, vol. 39, no. 3, pp. 667–702, 2014.

[16] T. Nguyen, D. Phung, B. Adams, and S. Venkatesh, "A sentiment-aware approach to community formation in social media," in *Proc. Int. AAAI Conf. Weblogs Soc. Media*, 2012, pp. 527–530.

[17] A. J. Rodriguez, S. E. Holleran, and M. R. Mehl, "Reading between the lines: The lay assessment of subclinical depression from written self-descriptions," *J. Personality*, vol. 78, no. 2, pp. 575–598, 2010.

[18] J. W. Pennebaker, M. E. Francis, and R. J. Booth, "*Linguistic Inquiry and Word Count (LIWC) [Computer Software]*," LIWC Inc, 2007.

[19] N. Ramirez-Esparza, C. K. Chung, E. Kacewicz, and J. W. Pennebaker, "The psychology of word use in depression forums in English and in Spanish: Testing two text analytic approaches," in *Proc. AAAI Int. Conf. Weblogs Soc. Media*, 2008, pp. 102–108.

[20] S. W. Stirman and J. W. Pennebaker, "Word use in the poetry of suicidal and nonsuicidal poets," *Psychosomatic Med.*, vol. 63, no. 4, pp. 517–522, 2001.

[21] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, 2003.

[22] M. J. Paul and M. Dredze, "Discovering health topics in social media using topic models," *PLoS ONE*, vol. 9, no. 8, 2014, Art. no. e103408.

[23] A. Dine, D. Larocque, and F. Bellavance, "Multivariate trees for mixed outcomes," *Comput. Statist. Data Anal.*, vol. 53, no. 11, pp. 3795–3804, 2009.

[24] N. Ghamrawi and A. McCallum, "Collective multi-label classification," in *Proc. ACM Int. Conf. Inform. Knowl. Manag.*, 2005, pp. 195–200.

[25] P. Rai, A. Kumar, and H. Daumé III, "Simultaneously leveraging output and task structures for multiple-output regression," in *Proc. Annu. Conf. Neural Inform. Process. Syst.*, 2012, pp. 3194–3202.

[26] T. L. Griffiths and M. Steyvers, "Finding scientific topics," *Proc. Nat. Acad. Sci. USA*, vol. 101, no. 90001, pp. 5228–5235, 2004.

[27] Y. Xu and W. Yin, "A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion," *SIAM J. Imaging Sci.*, vol. 6, no. 3, pp. 1758–1789, 2013.

[28] J. Liu, J. Chen, and J. Ye, "Large-scale sparse logistic regression," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2009, pp. 547–556.

[29] J. Liu, S. Ji, and J. Ye, "Multi-task feature learning via efficient l2, 1-norm minimization," in *Proc. Conf. Uncertainty Artif. Intell.*, 2009, pp. 339–348.

Authors' photographs and biographies not available at the time of publication.