A Computational Approach to Understand Mental Health from Reddit: Knowledge-Aware Multitask Learning Framework

Usha Lokala[1], Aseem Srivastava[2], Triyasha Ghosh Dastidar[3], Tanmoy Chakraborty[2], Md Shad Akthar[2], Maryam Panahiazar[4], Amit Sheth[1]

[1]Artificial Intelligence Institute, University of South Carolina, USA
[2]IIIT Delhi, India
[3]Dept. of CSE, BITS Pilani, India
[4]University of California, USA
nlokala@email.sc.edu, {aseems,shad.akhtar,tanmoy}@iiitd.ac.in, f20170829@hyderabad.bits-pilani.ac.in,
maryam.panahiazar@ucsf.edu, amit@sc.edu

Abstract

Analyzing gender is critical to study mental health (MH) support in CVD (cardiovascular disease). The existing studies on using social media for extracting MH symptoms consider symptom detection and tend to ignore user context, disease, or gender. The current study aims to design and evaluate a system to capture how MH symptoms associated with CVD are expressed differently with the gender on social media. We observe that the reliable detection of MH symptoms expressed by persons with heart disease in user posts is challenging because of the co-existence of (dis)similar MH symptoms in one post and due to variation in the description of symptoms based on gender. We collect a corpus of $150k$ items (both posts and comments) annotated using the subreddit labels and transfer learning approaches. We propose GeM, a novel task-adaptive multi-task learning approach to identify the MH symptoms in CVD patients based on gender. Specifically, we adapt a knowledge-assisted RoBERTa based bi-encoder model to capture CVD-related MH symptoms. Moreover, it enhances the reliability for differentiating the gender language in MH symptoms when compared to the state-of-art language models. Our model achieves high (statistically significant) performance and predicts four labels of MH issues and two gender labels, which outperforms RoBERTa, improving the recall by 2.14% on the symptom identification task and by 2.55% on the gender identification task.

Introduction

Social media platforms such as Twitter, TalkLife, and Reddit are extremely popular to express someone's feelings and views. In the past few years, people are using them extensively to share their mental and other general health issues. Cardiovascular Disease (CVD) is one such issue and is one of the leading causes of mortality in the United States and globally[1]. Gender, ethnic, racial, and age discrepancies within CVD diagnosis and treatment exist and have been well-reported (Kanchi et al. 2018). The activation of stress

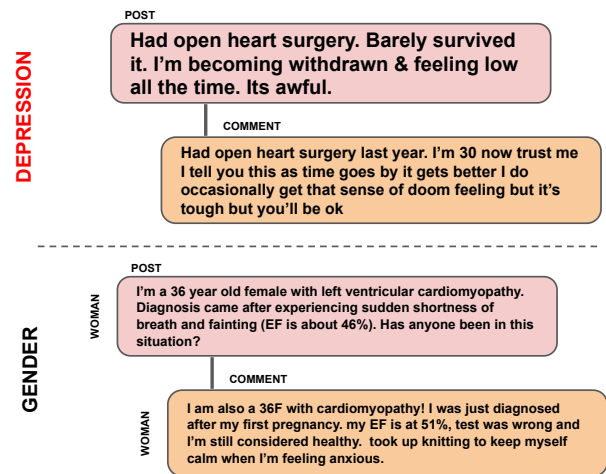[1]https://tinyurl.com/yh7yde4f

Figure 1: Our framework of social media posts contains four CVD MH symptoms– Depression, Anxiety, Bipolar, PTSD. We differentiate between gender for these communications. Our computational approach simultaneously identifies these symptoms and the underlying gender.

pathways in people leads to the mechanism that links Mental Health (MH) disorders and CVD. The American Journal of Medicine study established statistically significant links between MH disorders of depression, anxiety, psychological distress, and CVD (De Hert, Detraux, and Vancampfort 2018). They also found that the link is stronger in the case of depression-related symptoms and it can worsen the symptoms of CVD (Daniels et al. 2017). An article by the Cleveland Clinic also shows a particularly strong association between depression and CVD[2]. Thus, MH disorders are now established as risk factors for developing CVD, similar to the classic risk factors such as hypertension, diabetes mellitus, substance abuse, etc. Another dimension along which CVD can be analyzed is a person's gender.

Even though the general CVD risk factors in men and

[2]https://tinyurl.com/n2spyx2j

women are similar, the most common CVD symptoms and when and how they manifest vary significantly in women compared to men. For example, women are more likely to experience shortness of breath, nausea, backache, and vomiting while that is not the case for men[3]. As we see, a lot is known about the individual conditions and conditions in pairs such as CVD and MH, and CVD and gender-specific analysis; however, to the best of our knowledge, CVD has not been studied with both MH and gender-related risk factors. Obtaining clinical data for such a task is necessary to practice safe and ethical standards; however, it is highly cumbersome because it requires several expensive procedures and introduces a considerable time lag in procurement (Kilbridge 2003). In the digital era, it is well known that plenty of domain-specific knowledge on CVD and various MH disorders from reliable sources exist (Chaddha et al. 2016; Gkotsis et al. 2016; Ladeiras-Lopes et al. 2020b). However, the recent pandemic and overwhelmed healthcare system have pushed people to flock to online community platforms such as Reddit to post their concerns about gender-related issues on subreddits such as *r/askmen* and *r/askwomen*. This is possibly explained by Reddit potentially being one of the Internet's popular support groups (Park, Conway, and Chen 2018; Low et al. 2020). Though social media has been employed in past research for analysis of individual cases concerning CVD, the majority of it is focused more on Facebook, and Twitter (Ladeiras-Lopes et al. 2020a). The anonymity afforded on Reddit causes the posts to be more honest and less distorted by social perception and, thus, of sufficient quality to perform an investigative study. As millions of people use social media platforms for seeking MH support, understanding the symptoms varying with gender is crucial for web-based intervention. However, clinicians and health professionals are more interested in understanding MH disorders connected with the disease (Dhar and Barton 2016).

Problem Statement: In this paper, we seek to investigate the gender based MH detection among Reddit users afflicted by CVD. Specifically, we explore the question – *can gender detection improve the identification of MH in those with heart disease?* We observed that plenty of individuals openly discuss their MH symptoms on Reddit (Gaur et al. 2018; Gkotsis et al. 2016; Yadav et al. 2021; Kursuncu et al. 2019, 2018). However, a CVD MH-related post puts forward the issue of how each gender describes the symptom. In this work, we limit this study to two genders (reason being data inadequate for other genders for this particular study on social media), 'man' and 'woman', since 'man' and 'woman' include people who may or may not have been assigned 'male' or 'female' at birth. The reason these terms are more appropriate for this work is that, given the source of the data (Reddit is solely self-report), there is no way that we can know for each post whether they are man, woman, cis or trans at the time of birth, so it is more appropriate to use man/woman identity categories as opposed to male/female biological categorizations. We refer Our framework of social media posts and comments from Reddit (see Figure 1) contains four CVD MH symptoms and two gender labels for these communications.

---

[3]https://tinyurl.com/593hx3ae

Formally, we define the two tasks as follows: Let $S_i = s_{i1}, ..., s_{it}$ be the Reddit MH text (post/comment) and $G_i = g_{i1}, ..., g_{it}$ be the gender text (post/comment). For each S_i and G_i, two tasks are performed.

Task 1: Symptom Identification (SI) in CVD posts. The task is to identify the MH symptom, the post is talking about in the context of CVD. For each post/comment S_i in our dataset, we identify one of the four symptoms – *depression*, *anxiety*, *bipolar*, or *PTSD*.

Task 2: Gender Identification (GI) in CVD posts. In this task, we aim to detect the gender (man/woman) of the person showing CVD-related MH symptoms. This is to note that the post may not always be written by the user himself/herself; instead, in some cases, the post can be written by a close relative or a friend as well. For example, the following post, *She needs a cardiac procedure, which she is refusing to have. The cardiologist told her she can guarantee her she'll have a heart attack without it but still insists she's not having the procedure.* is originally posted by a 'man'. With so many woman-specific keywords, we want our model to predict gender correctly irrespective of the gender masking and writer of the post.

Proposed Method and Contributions: We compile a high-quality CVD dataset (CVDS) by crawling six subreddits. The CVDS dataset consists of a total of $16k$ posts and comments extracted from the crawled $150k$ Reddit data points. The filtering procedure applied to extract $16k$ posts and comments is described in Section 3. Moreover, the annotated dataset is generated using the automatic labelling methods and is validated by the domain experts. We develop a computational approach to understand the gender-based MH symptoms on social media. Our approach, which is knowledge aware, task adaptive RoBERTa based bi-encoder model, simultaneously identifies these symptoms and the underlying gender. We summarize our contributions below:

1. We propose a multi-task learning (MTL) framework for identifying MH symptoms that consider gender language fused in the CVD-related user posts. This is the first study in the CVD domain that considers gender language in MH-related posts to the best of our knowledge.

2. To justify the variations in the gender language description of symptoms, we propose a **G**ender **E**nabled **M**H framework, `GeM`, which is a RoBERTa-based bi-encoder model. The performance of `GeM` is improved for the task of SI along with the task of GI in a MTL paradigm.

3. To evaluate our model, we created a corpus of $150k$ data points with $16k$ posts and comments annotated (CVDS) with four MH symptoms. Also, these posts were labeled with the gender classes – man and woman. Resources created as a part of the study will be made available upon request to the corresponding author.

Related Work

MH Setting: Due to the superior performance of deep learning algorithms in data-rich healthcare applications (Miotto et al. 2018), computational methods to model MH outcomes

using deep learning are fast gaining prominence. Shatte, Hutchinson, and Teague explored the application of machine learning techniques in MH for diagnosis, prognosis, and treatment. Pruksachatkun, Pendse, and Sharma used machine learning to identify changes in MH behavior using topic and sentiment analysis of the posts. Barney, Griffiths, and Banfield modeled depression concerns of people. Chancellor and De Choudhury interpreted MH based on social media knowledge. Gaur et al.; Gaur et al. identified MH needs of people posting on online Reddit communities using deep learning methods based on time-variant and time-invariant analyses. Malhotra et al. showed transformer based architectures to analyze MH using conversational data from social media apps and online domain knowledge. Dolbir, Dastidar, and Roy used transformer models such as GPT-2 to understand the MH condition of the user post. As we can see, there is a large body of work in utilizing deep learning for MH prediction and analysis (Akbulut, Şengür, and Ekici 2017). To understand the effect of gender on MH, (De Choudhury et al. 2013) considered self reported gender information, symptoms via social media to predict depression. (Lindsey, Joe, and Nebbitt 2010) examined the influence of mental health stigma on depressive symptoms which considered gender as an important role on how the symptoms are described.

CVD Setting: Mathur et al. surveyed techniques to predict CVD outcomes, risk factors, early rule-based symptoms limited the use of computational technology for clinical decision-making. Thus, following the advent of deep learning, a wide range of CVD applications, including prediction of drug therapy, pharmacogenomics, heart failure management, cardiovascular imaging, and diagnostics, has been developed (Kalinin et al. 2018; Li et al. 2020; Powles and Hodson 2017). Social media analysis on prediction of CVD-related symptoms, risk factors, behaviors and outcomes, to our knowledge has not been studied before.

Gender Detection: Gender detection from various types of data including images and text on social media such as Facebook, Twitter and even from different language and dialects has been extensively studied using deep learning models (ElSayed and Farouk 2020). Akbulut, Şengür, and Ekici utilized knowledge from local receptive fields and CNNs to identify the gender of a person from their faces. Kowsari et al. used an ensemble method to identify the gender from social networks. ElSayed and Farouk identified gender from Egyptian Arabic dialect tweets using a variety of neural networks, including vanilla deep networks, CNNs, LSTM, GRU, both uni-directional and bi-directional. However, there is little work on gender detection from a resource-rich platform such as Reddit. Thus, we see that Reddit has been relatively less explored for CVD and gender detection even if MH research has been carried out on social media platform. Thus analysis of Reddit for a cross-sectional study of these MH symptoms and gender factors is a highly challenging and necessary research problem.Furthermore, to the best of our knowledge, though deep learning architectures have been developed in healthcare for CVD, MH, gender-based studies as well as multimodal data, no prior work has investigated a combination of all three factors.

	Posts	Comments	Users (P)	Users(C)
Depression	2125	2281	1966	1911
Anxiety	1983	2498	1635	1804
Bipolar	1796	2016	1044	1128
PTSD	1831	2027	1155	969
man	3554	4637	2645	3225
woman	4181	4185	3014	2358

Table 1: Dataset statistics on the CVDS corpus used in our experiments for SI and GI tasks. (P) and (C) represent the set of users who have posted and commented, respectively.

	Posts	Comments	Users (P+C)
r/AskMen	13659	22706	8347
r/AskWomen	14823	21673	7955

Table 2: Dataset statistics on the gender knowledge base corpus used in transfer learning to annotate data with gender labels. (P) and (C) represent the set of users who have posted and commented, respectively.

The CVDS Dataset Development

In this section, we describe our effort in developing the CVD dataset using Reddit. We also describe our annotation procedure for the same. We use the entire dataset for task adaptive pre-training and annotate a subset of them containing $8k$ posts and $8k$ comments.

Data Collection: We collect posts and comments from MH-related subreddits for this study. We consider four MH subreddits – r/depression, r/anxiety, r/PTSD, and r/bipolar. We choose only these subreddits as these are the major CVD symptoms related to MH as suggested in (CDC 2021; Ormel et al. 2007). Apart from this, we collect data from two gender-related subreddits – r/AskMen and r/AskWomen. The Reddit API was used to crawl six subreddits for a total of $150k$ posts and comments.

Identifying CVD Discussions: We crawl the data using a carefully curated CVD lexicon consisting of around 250 CVD-related terms obtained from various sources like SNOMED-CT[4], DataMed[5] and ICD-10[6]. Moreover, we append the list with commonly used CVD-related topics and abbreviations such as CABG (Coronary artery bypass graft), CAD (Coronary artery disease), CHD (Coronary heart disease), and CHF (Congestive heart failure). We collect these CVD topics and abbreviations from Texas Heart Institute (Tex 2017). Utilizing the compiled list, we obtain $\sim 16K$ posts and comments from $5,400$ users. The CVDS data stats are presented in Table 1 with number of posts and comments for each MH symptom and Gender and number of users who posted and commented.

Automatic Labelling and Annotations: The posts in these subreddits are assigned a class label corresponding to the name of the MH disorder they are associated with. We consider posts with a number of upvotes of more than 10

---

[4]https://tinyurl.com/nczubnzr

[5]https://datamed.org/

[6]https://tinyurl.com/kyky45kz

TL-BERT Model	Precision	Recall	F1
Gender Post	87.36	83.47	85.92
Gender Comment	78.78	79.32	78.90

Table 3: Validation results for GI through the transfer learning approach on generic (non-CVD) related posts. The trained model is then used to obtain the gender labels for the posts and comments in the CVDS dataset.

and a minimum token length of 50 to maintain the quality. We further remove any URLs or usernames that could potentially contain sensitive information. For generating gender labels, we implement an inductive transfer learning approach with BERT (Mozafari, Farahbakhsh, and Crespi 2019). For this, we additionally crawl $72k$ posts from r/AskMen and r/AskWomen. We observed that 92% of the posts posted in r/askmen are posted by man and the 94% of posts from r/askwomen are posted by woman. We split this dataset into train, dev, and test sets (75:5:20). We train a BERT-based text classifier for 10 epochs using learning rate of 1e-5, batch size of 32 to distinguish the posts written by man and woman. Subsequently, gender labels of $16k$ posts are fetched in the CVDS dataset. We also examine manual inter-annotator agreement among three domain experts for gender and symptom labels of 300 posts and comments which is 0.72 Kappa score with biomedical informatics specialist co-author to validate the annotations. The manual annotations are evaluated in same way as automated labels and our macro F measure against ground truth is 0.70. The stats of the posts and comments used for training the transfer learning model are found in Table 2. The results for transfer learning using BERT are reported in Table 3. The statistics of the CVDS and gender datasets are shown in Tables 1 and 2, respectively.

Proposed Methodology

We develop a computational approach, GeM as shown in Figure 2 to understand the gender-specific MH symptoms in CVD. We formulate GeM as a model consisting of knowledge aware entity masking combined with the task adaptive RoBERTa model. Here, we describe the drug abuse ontology as one of the sources of knowledge and the other modules in our architecture.

Drug Abuse Ontology: The Drug Abuse Ontology (DAO) (Cameron et al. 2013; Lokala et al. 2020) is a domain-specific knowledge source containing drug and health-related classes, properties, relationships, and instances. Apart from medical terms, it includes concepts of MH disorders and symptoms aligned with the DSM-5[7] scale. Initially, it is used to determine user knowledge, attitudes, and behaviors related to non-medical use of buprenorphine and other illicit opioids through analysis of web forum data. Later, this ontology evolved to understand MH disorders and trends of cannabis product use in the context of changing cannabis legalization

policies in the USA. Recently, this proved effective in capturing gleaning trends in the availability of novel synthetic opioids through analysis of crypto market data. In this study, we leverage DAO to map the symptoms we find on social media to the parent concepts for MH conditions and associated symptoms. For example, properties for '*anxiety*' include terms such as 'feeling jittery', 'antsy', 'feeling doomed', 'nausea', 'lack of concentration', and 'throwing up'.

Entity Masking with External Knowledge:

Symptom Masking: Extracting entities and relationships related to MH symptoms and gender-like concepts/synonyms/slang terms requires domain knowledge. For a text S_i, the concepts/symptoms/synonyms related to MH labels (depression, anxiety, PTSD, bipolar) are masked using the MedDRA Ontology[8] and DAO. We use these ontologies to map entities and n-grams in a post/comment to their parent concept in the ontology. We further leverage the potential of detecting clinical symptoms in post/comment as per the clinical questionnaires like PHQ-9 (Kroenke, Spitzer, and Williams 2001) for Depression, GAD-7 (Jordan, Shedden-Mora, and Löwe 2017) for Anxiety, CAPS-5 (Weathers et al. 2018) for PTSD, and BDRS (Berk et al. 2007) for Bipolar to perform masking (Yazdavar et al. 2017; Yadav et al. 2021). This approach provides the model with more context about the main entities of interest in the text. It removes noise in the data, thus contributing to the robustness of the model without overfitting that may happen if the model tries to learn noise in data. Our ablation study (Table 9) shows that external knowledge adds value to data in the model decision-making. The reason for masking is not to bias the model to specific entities. For example, the entities or symptoms related to PTSD in the post and comment are masked with <PTSD>; similarly, entities related to anxiety are masked with <anxiety>, and so on for the other three labels.

Gender Masking: We observe that the terms like *'F'* and *'M'* on subreddits indicate woman and man, respectively. Similarly, there are many other terms that indicate gender indirectly. For gender word masking, we use 'Gendered words'[9] lexicon which is created from the WordNet synset. Each word in this dataset is manually tagged as gender-neutral (n), man or masculine (m), woman or feminine (f), or other (o). However, for this study, we only use 'man' and 'woman' tags. Words that cannot be extracted from WordNet are manually added, including personal (he, she, him, her) and possessive (his, hers, her) pronouns. The words related to gender in the post/comments are masked with either <man> or <woman>. For example, bachelorette is masked as <woman> and bachelor is masked as <man>. The input to the model is a Reddit post or comment, and the output is the MH symptom and gender.

Knowledge Assisted Bi-encoder Model: We propose a knowledge assisted bi-encoder model based on RoBERTa (Liu et al. 2019) for identifying symptom and gender as shown in Figure 2. We multitask over the two tasks of SI and GI with data in CVDS. We observe that bi-encoder architecture with external knowledge and attention provide more

---

[7]Diagnostic and Statistical Manual of Mental Disorders - 5th Edition

[8]https://tinyurl.com/s2upzusc

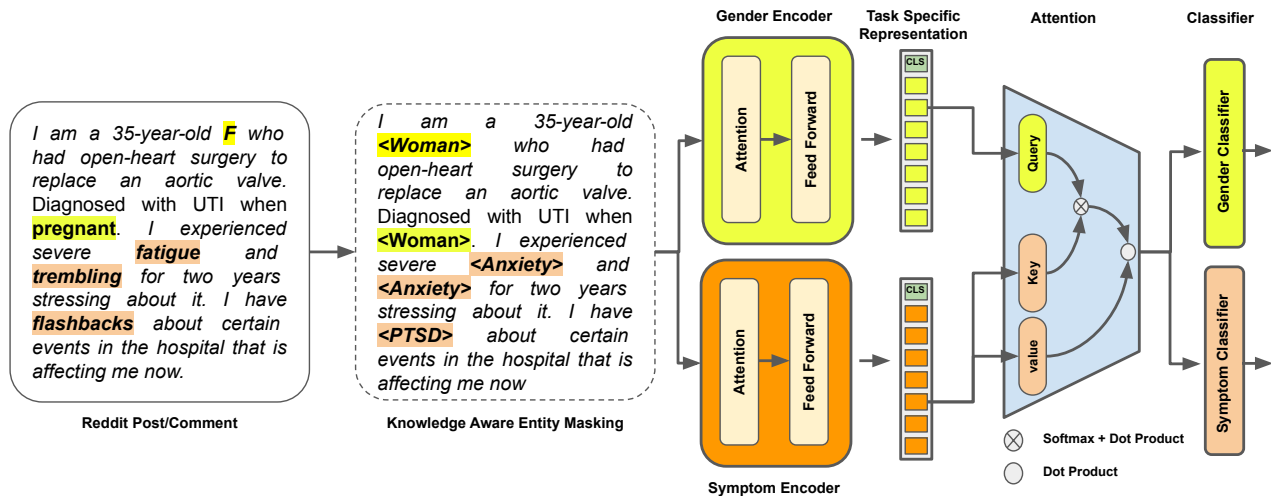[9]https://github.com/ecmonsen/gendered_words

Figure 2: Proposed model architecture of **GeM** for SI and GI tasks.

context than working with the encoder itself. The two encoders Symptom-encoder (S-encoder) and Gender-encoder (G-encoder) for symptom and gender, respectively, are pretrained on RoBERTa. The S-encoder is for understanding the context of MH symptoms, whereas, the G-encoder is for understanding the gender-related language in the post/comment.

Task Adaptive Pretraining: Both S-encoder and G-encoder are initialized using the weights learned by RoBERTa$_{BASE}$. We further perform a task-adaptive pretraining (Gururangan et al. 2020) of the two encoders to adapt to the symptom and gender-specific context. For this additional pre-training of the two encoders, we use the datasets of $65k$ posts and $150k$ comments from several MH subreddits (bipolar, anxiety, depression, and PTSD) and gender-related subreddits (r/AskMen, r/AskWomen). For this task, the masked language modeling task for pre-training (10 epochs, batch size 32) is used.

Attention Layer: Subsequently, we compute attention over the two encodings for generating MH context-aware representation of gender text. Our query is the gender text encoding $e_i^{(G)}$, while the key and value are the symptom text encoding $e_i^{(S)}$. We compute the attention as:

$$ a_i(e_i^{(S)}, e_i^{(G)}) = softmax(\frac{e_i^{(S)} e_i^{(G)}}{\sqrt{d}}) e_i^{(G)} $$

where $d = 768$ is the hidden dimension in RoBERTa$_{BASE}$. Next, we add the encoded gender post $e_i^{(G)}$ with its representation transformed through attention $a_i(e_i^{(S)}, e_i^{(G)})$ to extract a residual mapping, $h_i^{(G)}$ (He et al. 2016). It forms the final MH context-aware representation of the input post.

Prediction Layer: Finally, we take the final representations of the [CLS] tokens and pass it through two separate linear layers for the gender and symptoms predictions.

Model	SI-Posts			SI-Comment		
	P	R	F1	P	R	F1
LSTM$_{STL}$	71.67	70.15	70.90	75.51	75.53	75.51
BERT$_{STL}$	76.35	76.33	76.33	76.53	76.33	76.42
GPT-2$_{STL}$	76.23	76.85	76.53	76.79	76.60	76.69
RoBERTa$_{STL}$	82.73	82.41	82.56	80.10	81.28	80.68
BERT$_{MTL}$	75.24	77.12	76.16	74.65	77.45	76.53
RoBERTa$_{MTL}$	81.34	83.15	82.23	81.25	83.74	82.47
GeM	**84.18**	**85.29**	**84.73**	**83.57**	**84.07**	**83.81**

Table 4: Performance comparison our proposed model with the baselines for SI on both the posts and comments. SI: Symptom Identification, P: Precision, R:Recall.

Experiments and Results

In this section, we present our experimental results in identifying MH symptoms with underlying gender. We compare the performance of GeM against state-of-the-art models used in similar tasks. Moreover, we also present our prediction analysis and how challenging it is to predict MH symptoms based on gender.

Baseline: (1) LSTM (Hochreiter and Schmidhuber 1997); (2) BERT (Devlin et al. 2018): This is a BERT base model which is finetuned for individual tasks as single-task learning (STL) for MH and gender detection. We also implement BERT in the MTL setting for the tasks of SI and GI. (3) GPT-2 (Radford et al. 2019). (4) RoBERTa$_{BASE}$ (Liu et al. 2019).

Experimental Setup: For experiments, we split the CVDS dataset into 75:5:20 ratio for the train, development, and test sets, respectively. We fine tune the hyper-parameters using the development set. Each model is trained for 10 epochs with a learning rate of 1e-5, batch size of 64. We use cross-entropy loss and ADAM (Kingma and Ba 2014) for the optimization. For regularization, we use dropout (Hinton et al. 2012) with a probability of 0.2. For both tasks, we report recall, precision, and F1-score.

Results: Tables 4 and 5 presents the results of GeM with ex-

isting models, indicating that GeM significantly outperformed the previous architectures. Tables 4 and 5 report the performance for the SI and GI tasks, respectively. Table 6 presents the class-wise results of GeM for symptom and gender on posts and comments. LSTM performs the worst among the four baselines while RoBERTa achieves the best performance. However, GeM outperforms RoBERTa with substantial gains demonstrating the performance of our 'knowledge aware bi-encoder' model over other baselines for this task. Basically, we train our baselines in two different ways – STL and MTL. We can see in Table 4 and Table 5 that MTL framework performs well in the main task (MH symptom detection) and also the secondary task (gender detection). However, BERT and GPT-2 perform similarly in the STL setting while RoBERTa performed better both for main task in the STL setting and in MTL setting. But for the gender detection task, BERT performs better than GPT-2 in the STL setting while RoBERTa performs similarly in STL and MTL settings. This demonstrates that providing an additional attention mechanism improves the performance of GeM over the strong self-attention mechanism of RoBERTa for this task. To get a deeper understanding of how GeM performed over the baseline models, we observe the following on the classification of posts and comments on SI and GI tasks.

- **Understanding MH symptom:** Table 7 shows the capability of our model to capture symptoms like depression, anxiety, bipolar and PTSD. Our model performs better than GPT-2 and LSTM in predicting Depression and PTSD. The reason why BERT misclassified Bipolar and Anxiety maybe because of BERT model trained on sources like Wikipedia, Books compared to MH symptoms described in social media. GeM is able to detect classes-bipolar and PTSD with high accuracy among posts and comments compared to classes-depression and anxiety (Table 6). Even with just the posts, GeM is able to predict the Biploar class with an F1 score of 87%.

- **Understanding Gender:** For the posts where gender was not explicitly mentioned, most of the baselines are prone to a false classification as shown in Table 7. BERT fails to model the meaning of gender context inside post. If the post talks about his/her family, some of the baselines fail to understand the primary gender the post talks about. However, our model with the MH aware attention information sharing unit tends to have better coverage for identifying the symptoms related to CVD in predicting the primary gender the post is meant to talk about. GeM can detect gender classes with high accuracy among both posts and comments relatively well (Table 6).

Error Analysis

We qualitatively analyze the sources of our errors (Table 8). The following are the major errors made by our approach.

1. **Co-existing Symptoms:** GeM is not able to predict correctly if multiple symptoms co-exist, e.g., when depression might often occur with another MH symptom in a post (c.f. Table 8).

2. **Class-level Ambiguity:** Regarding the class-wise performance among MH symptoms, the best performing classes

Model	GI-Posts			GI-Comment		
	P	R	F1	P	R	F1
LSTM$_{STL}$	64.95	65.11	65.02	70.99	70.78	70.88
BERT$_{STL}$	78.78	79.32	79.04	76.53	76.33	76.42
GPT-2$_{STL}$	70.20	70.75	70.47	72.57	71.27	71.91
RoBERTa$_{STL}$	78.80	78.34	78.56	80.53	79.42	79.97
BERT$_{MTL}$	77.65	75.34	76.47	77.54	74.38	75.92
RoBERTa$_{MTL}$	79.32	77.32	78.30	79.47	77.65	78.54
GeM	**80.34**	**79.87**	**80.10**	**80.53**	**79.67**	**80.09**

Table 5: Performance comparison our proposed model with the baselines for GI on both the posts and comments. GI: Gender identification, P: Precision, R:Recall.

Class	Posts			Comments		
	P	R	F1	P	R	F1
Depression	82.71	84.24	83.46	78.45	80.63	79.52
Anxiety	80.28	84.59	82.37	79.26	81.81	80.51
Bipolar	88.87	85.36	87.07	87.28	87.46	87.36
PTSD	84.66	86.97	85.79	89.26	86.38	87.79
man	79.37	77.26	78.30	80.78	76.15	78.39
woman	81.32	82.51	81.91	80.29	83.18	81.70

Table 6: Class-wise performance of GeM for MH symptoms and gender on both the posts and comments. P: Precision, R:Recall.

are PTSD and bipolar, while the worst performing classes are depression and anxiety (Table 6). The reason can be attributed to the word depression occurs in 23% of total posts related to other symptoms, and anxiety occurs in 31% of total posts related to other symptoms. This implies that the model cannot capture such class mentions as it does with other MH symptom classes. This may explain the relatively lower precision for depression and anxiety.

3. **Ambiguity in Gender:** Posts mentioning about the persons themselves are easy to classify in gender than a post talking about others and then transitioning back to self-description (Table 8). However, mentions in posts such as 'my mother' or 'my brother' are correctly predicted in gender as shown in Table 7. Our model is able to achieve higher performance for 'woman' than 'man' both for posts and comments as they are described with more communication, self representation (Haferkamp et al. 2012) and self declared gender descriptions which attributed to better performance.

Ablation Study

To determine statistical significance, we perform Wilcoxon Signed Rank Test (Woolson 2007) on MH symptom expressed in posts and comments between where gender information is present and absent. We observe that the cosine similarity between MH symptom expressed by post/comment with gender present is significantly ($p < 0.001$) lower than a gender absent post/comment. We next analyze the components and training strategies in our approach through an ablation study by removing one component from and evaluating the performance. To remove the attention component

Post/Comment	Case	Actual	Predicted			
			BERT$_{MTL}$	**GPT-2**	**RoBERTa**$_{MTL}$	**GeM**
I'm having heart surgery in 2 weeks. This is too much for someone who has been suffering from panic attacks	Symptom	Anxiety	PTSD	Bipolar	Anxiety	Anxiety
I've noticed that a lot of my CVD symptoms have escalated over time. My flash backs have become much more 'vivid', the adrenaline response to panic has increased, my libido has plummeted.		PTSD	PTSD	Depression	Anxiety	PTSD
I was on Prozac for a while, but I was still having panic attacks once or twice a month. I think mine are related to my menstrual hormones	Gender	woman	man	woman	man	woman
My grandma had a stress-induced heart attack 4 years ago. My grandpa, her husband of over 60 years, had dementia, and she was his sole carer		woman	man	man	woman	woman

Table 7: Qualitative analysis of proposed approach GEM, with the baseline models.

Post/Comment	Error Type	Actual	Predicted
Issues with IUD in uterus since the first heart attack. My grandfather didn't get stronger too after he survived his heart attack. It left him weak and he spent the rest of his days wishing he died. if anything it's left me with a mountain of debt and a deep depression.	Ambiguity in Gender	woman	man
I get panic attacks and depression. The first time I had a panic attack I was driving on the highway and thought that I was going to seriously have a heart attack and die. It was mixture of fear and panic in one.	Co-existing Symptom	Anxiety	Depression
My ptsd came about from a traumatic experience of terror but my depression and anxiety plummeted. They all experienced high anxiety problems too. The other two ended up with serious heart problems that forced them into surgery and unemployment. We all ended up poor. I ended up homeless.	Class-level Ambiguity	PTSD	Anxiety

Table 8: Examples showing major errors made by our proposed approach.

Model	Precision	Recall	F1
GeM	**84.18**	**85.29**	**84.73**
-Attention (A)	81.22 (2.96↓)	82.13 (3.16↓)	81.67 (3.06↓)
-EM	79.34 (4.84↓)	80.68 (4.61↓)	80.00 (4.73↓)
-TA	83.25 (0.93↓)	84.13 (1.16↓)	83.68 (1.05↓)

Table 9: Ablation Study: Median of metrics over 10 different runs. Bold denotes best performance. EM: Entity Masking, TA: Task Adaptation

from the model, instead of using attention, we concatenate the symptom post encoding $e_i^{(S)}$ with the Gender post encoding $e_i^{(G)}$ and use the concatenated representation as input to the linear layer. For removing the entity masking component, we train our encoders with raw data directly extracted from the social media. Task adaptive pre-training is removed from the model by initializing the model weights from RoBERTa. We present the results on SI task for posts in Table 9. Our major gains come from using entity masking which signifi-

cantly benefits the SI task (+2.96 precision, +3.16 recall). The result is significantly better than model without entity masking ($p < 0.05$) for Wilcoxon Signed Rank test which shows that contextualized representation is highly desirable for the SI task in this study. Also, using task adaptive pre-training only leads to minor performance improvement. We further observe that attention improved the precision of the model by 3.16% making every aspect of the model contributing to the performance of this task.

Potential Impact and Conclusion

In this paper we explored a new health dimension of social media in CVD to identify MH symptoms. We created a domain-specific dataset (CVDS) for identifying MH symptoms with underlying gender. We developed a novel framework and a computational method GeM for understanding gender language in MH-based text and asynchronous conversations on MH platforms. We introduced a task adaptive

RoBERTa-based attention framework that learns to identify the MH symptoms with the underlying task of gender detection. Our computational approach effectively identified gender while learning the MH symptoms in the CVD context. Our experimental results show the effectiveness of studying gender-based MH symptoms in a specific disease context. The limitations in this study are only two genders and four CVD related MH symptoms are considered. In the future, we aim to extend our Reddit dataset to other social media platforms, genders, MH disorders and the dataset can potentially grow to be a gold standard on identifying MH symptoms for people with CVD. We will further test our model for EMR data to better understand the gains and pitfalls in actual clinical settings. Through GeM, we hope to form a future component to process a larger EMR data set to (a) validating the new clinical process guidelines for patients with CVD and creating gender and sex-based Order Sets and updated questionnaires in EMR and (b) creating actionable gender-based order Sets integrated into the EMR system that will help the providers with decision-making and avoid possible delays and discrepancies in the Diagnosis and Treatment of CVD.

Ethical Statement

We apply our model to study how gender impacts online MH conversation dynamics related to CVD. An important aspect that we need to consider while working with health-related issues is to respect the users' privacy and adhere to good ethical practices adapted by previous research (Henderson, Johnson, and Auld 2013; Alim 2014). Therefore, similar to Matthews2017-cn, we censor several sensitive information, such as user names, personal information, platform identifiers, and urls which might be directly linked to the users' identity, from the collected posts and comments. All examples used in this paper are anonymized, and de-identified for user privacy (Hunter et al. 2018). We also adopted the proposed guidelines for treatment of Names and Online Pseudonyms in posts and comments gathered from Reddit (Bruckman 2002).

In this work, We study expressed symptoms in mental health subreddit groups in the form of textual interactions. The expressed symptom may differ from the symptom actually perceived or experienced by the person. However, obtaining perceived symptoms from social media is challenging and involves ethical risks. We can adopt the approach used in this work to develop the data set with more human supervision and we acknowledge that the data may be prone to demographic, annotator, and platform-specific biases (Preoţiuc-Pietro et al. 2015; Wang et al. 2019). We also acknowledge that the current work does not make any clinical diagnosis or treatment suggestions in any manner whatsoever.

Acknowledgments

References

2017. Heart Information Center. https://www.texasheart.org/heart-health/heart-information-center/. Accessed: 2021-6-12.

Akbulut, Y.; Şengür, A.; and Ekici, S. 2017. Gender recognition from face images with deep learning. In *2017 International artificial intelligence and data processing symposium (IDAP)*, 1–4. IEEE.

Alim, S. 2014. An initial exploration of ethical research practices regarding automated data extraction from online social media user profiles. *First Monday* .

Barney, L. J.; Griffiths, K. M.; and Banfield, M. A. 2011. Explicit and implicit information needs of people with depression: a qualitative investigation of problems reported on an online depression support forum. *BMC psychiatry* 11(1): 1–11.

Berk, M.; Malhi, G. S.; Cahill, C.; Carman, A. C.; Hadzi-Pavlovic, D.; Hawkins, M. T.; Tohen, M.; and Mitchell, P. B. 2007. The Bipolar Depression Rating Scale (BDRS): its development, validation and utility. *Bipolar disorders* 9(6): 571–579.

Bruckman, A. 2002. Studying the amateur artist: A perspective on disguising data collected in human subjects research on the Internet. *Ethics and Information Technology* 4(3): 217–231.

Cameron, D.; Smith, G. A.; Daniulaityte, R.; Sheth, A. P.; Dave, D.; Chen, L.; Anand, G.; Carlson, R.; Watkins, K. Z.; and Falck, R. 2013. PREDOSE: a semantic web platform for drug abuse epidemiology using social media. *Journal of biomedical informatics* 46(6): 985–997.

CDC. 2021. Heart disease and mental health disorders. https://www.cdc.gov/heartdisease/mentalhealth.htm. Accessed: 2021-5-26.

Chaddha, A.; Robinson, E. A.; Kline-Rogers, E.; Alexandris-Souphis, T.; and Rubenfire, M. 2016. Mental Health and Cardiovascular Disease. *Am. J. Med.* 129(11): 1145–1148.

Chancellor, S.; and De Choudhury, M. 2020. Methods in predictive techniques for mental health status on social media: a critical review. *NPJ digital medicine* 3(1): 1–11.

Daniels, A.; et al. 2017. Mental health and cardiovascular disease. *Australian Pharmacist* 36(7): 1.

De Choudhury, M.; Gamon, M.; Counts, S.; and Horvitz, E. 2013. Predicting depression via social media. In *Seventh international AAAI conference on weblogs and social media*.

De Hert, M.; Detraux, J.; and Vancampfort, D. 2018. The intriguing relationship between coronary heart disease and mental disorders. *Dialogues in clinical neuroscience* 20(1): 31.

Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. doi:10.48550/ARXIV.1810.04805.

Dhar, A. K.; and Barton, D. A. 2016. Depression and the link with cardiovascular disease. *Frontiers in psychiatry* 7: 33.

Dolbir, N.; Dastidar, T.; and Roy, K. 2021. NLP is Not enough–Contextualization of User Input in Chatbots. *arXiv preprint arXiv:2105.06511* .

ElSayed, S.; and Farouk, M. 2020. Gender identification for Egyptian Arabic dialect in twitter using deep learning models. *Egyptian Informatics Journal* 21(3): 159–167.

Gaur, M.; Aribandi, V.; Alambo, A.; Kursuncu, U.; Thirunarayan, K.; Beich, J.; Pathak, J.; and Sheth, A. 2021a. Characterization of Time-variant and Time-invariant Assessment of Suicidality on Reddit using C-SSRS. *Plos one* 16(5): e0250448.

Gaur, M.; Kursuncu, U.; Alambo, A.; Sheth, A.; Daniulaityte, R.; Thirunarayan, K.; and Pathak, J. 2018. "Let Me Tell You About Your Mental Health!": Contextualized Classification of Reddit Posts to DSM-5 for Web-based Intervention. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, CIKM '18, 753–762. New York, NY, USA: Association for Computing Machinery.

Gaur, M.; Roy, K.; Sharma, A.; Srivastava, B.; and Sheth, A. 2021b. " Who can help me?": Knowledge Infused Matching of Support Seekers and Support Providers during COVID-19 on Reddit. *arXiv preprint arXiv:2105.06398* .

Gkotsis, G.; Oellrich, A.; Hubbard, T.; Dobson, R.; Liakata, M.; Velupillai, S.; and Dutta, R. 2016. The language of mental health problems in social media. In *Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology*, 63–73. San Diego, CA, USA: Association for Computational Linguistics.

Gururangan, S.; Marasović, A.; Swayamdipta, S.; Lo, K.; Beltagy, I.; Downey, D.; and Smith, N. A. 2020. Don't Stop Pretraining: Adapt Language Models to Domains and Tasks .

Haferkamp, N.; Eimler, S. C.; Papadakis, A.-M.; and Kruck, J. V. 2012. Men are from Mars, women are from Venus? Examining gender differences in self-presentation on social networking sites. *Cyberpsychology, Behavior, and Social Networking* 15(2): 91–98.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

Henderson, M.; Johnson, N. F.; and Auld, G. 2013. Silences of ethical practice: dilemmas for researchers using social media. *Educ. Res. Eval.* 19(6): 546–560.

Hinton, G. E.; Srivastava, N.; Krizhevsky, A.; Sutskever, I.; and Salakhutdinov, R. R. 2012. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580* .

Hochreiter, S.; and Schmidhuber, J. 1997. Long short-term memory. *Neural computation* 9(8): 1735–1780.

Hunter, R. F.; Gough, A.; O'Kane, N.; McKeown, G.; Fitzpatrick, A.; Walker, T.; McKinley, M.; Lee, M.; and Kee, F. 2018. Ethical issues in social media research for public health. *American Journal of Public Health* 108(3): 343–348.

Jordan, P.; Shedden-Mora, M. C.; and Löwe, B. 2017. Psychometric analysis of the Generalized Anxiety Disorder scale (GAD-7) in primary care using modern item response theory. *PloS one* 12(8): e0182162.

Kalinin, A. A.; Higgins, G. A.; Reamaroon, N.; Soroushmehr, S.; Allyn-Feuer, A.; Dinov, I. D.; Najarian, K.; and Athey, B. D. 2018. Deep learning in pharmacogenomics: from gene regulation to patient stratification. *Pharmacogenomics* 19(7): 629–650.

Kanchi, R.; Perlman, S. E.; Chernov, C.; Wu, W.; Tabaei, B. P.; Trinh-Shevrin, C.; Islam, N.; Seixas, A.; Rodriguez-Lopez, J.; and Thorpe, L. E. 2018. Gender and race disparities in cardiovascular disease risk factors among New York City adults: New York City Health and Nutrition Examination Survey (NYC HANES) 2013–2014. *Journal of Urban Health* 95(6): 801–812.

Kilbridge, P. 2003. The cost of HIPAA compliance. *The New England journal of medicine* 348(15): 1423.

Kingma, D. P.; and Ba, J. 2014. Adam: A Method for Stochastic Optimization. *CoRR* abs/1412.6980.

Kowsari, K.; Heidarysafa, M.; Odukoya, T.; Potter, P.; Barnes, L. E.; and Brown, D. E. 2020. Gender detection on social networks using ensemble deep learning. In *Proceedings of the Future Technologies Conference*, 346–358. Springer.

Kroenke, K.; Spitzer, R. L.; and Williams, J. B. 2001. The PHQ-9: validity of a brief depression severity measure. *Journal of general internal medicine* 16(9): 606–613.

Kursuncu, U.; Gaur, M.; Lokala, U.; Illendula, A.; Thirunarayan, K.; Daniulaityte, R.; Sheth, A.; and Arpinar, I. B. 2018. What's ur Type? Contextualized Classification of User Types in Marijuana-Related Communications Using Compositional Multiview Embedding. In *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, 474–479.

Kursuncu, U.; Gaur, M.; Lokala, U.; Thirunarayan, K.; Sheth, A.; and Arpinar, I. B. 2019. Predictive Analysis on Twitter: Techniques and Applications. In Agarwal, N.; Dokoohaki, N.; and Tokdemir, S., eds., *Emerging Research Challenges and Opportunities in Computational Social Network Analysis and Mining*, 67–104. Cham: Springer International Publishing.

Ladeiras-Lopes, R.; Baciu, L.; Grapsa, J.; Sohaib, A.; Vidal-Perez, R.; Bohm, A.; Silvola, H.; Gimenez, M. R.; Muscoli, S.; Wallner, M.; et al. 2020a. Social media in cardiovascular medicine: a contemporary review. *European Heart Journal-Digital Health* 1(1): 10–19.

Ladeiras-Lopes, R.; the 'Cardiologists of Tomorrow', Digital Health and Media Committees of the European Society of Cardiology; Baciu, L.; Grapsa, J.; Sohaib, A.; Vidal-Perez, R.; Bohm, A.; Silvola, H.; Gimenez, M. R.; Muscoli, S.; Wallner, M.; Rakisheva, A.; Nagy, V. K.; Cowie, M. R.; Clarke, S. C.; and Achenbach, S. 2020b. Social media in cardiovascular medicine: a contemporary review. *Eur Heart J Digit Health* 1(1): 10–19.

Li, Q.; Wang, J.; Tao, H.; Zhou, Q.; Chen, J.; Fu, B.; Qin, W.; Li, D.; Hou, J.; Chen, J.; et al. 2020. The prediction model of warfarin individual maintenance dose for patients undergoing heart valve replacement, based on the back propagation neural network. *Clinical drug investigation* 40(1): 41–53.

Lindsey, M. A.; Joe, S.; and Nebbitt, V. 2010. Family matters: The role of mental health stigma and social support on depressive symptoms and subsequent help seeking among African American boys. *Journal of Black Psychology* 36(4): 458–482.

Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; and Stoyanov, V. 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach .

Lokala, U.; Daniulaityte, R.; Lamy, F.; Gaur, M.; Thirunarayan, K.; Kursuncu, U.; and Sheth, A. P. 2020. DAO: An Ontology for Substance Use Epidemiology on Social Media and Dark Web. *JMIR Public Health and Surveillance* .

Low, D. M.; Rumker, L.; Talkar, T.; Torous, J.; Cecchi, G.; and Ghosh, S. S. 2020. Natural language processing reveals vulnerable mental health support groups and heightened health anxiety on reddit during covid-19: Observational study. *Journal of medical Internet research* 22(10): e22635.

Malhotra, G.; Waheed, A.; Srivastava, A.; Akhtar, M. S.; and Chakraborty, T. 2022. Speaker and Time-Aware Joint Contextual Learning for Dialogue-Act Classification in Counselling Conversations. In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, WSDM '22, 735–745. New York, NY, USA: Association for Computing Machinery. ISBN 9781450391320. doi:10.1145/3488560.3498509. URL https://doi.org/10.1145/3488560.3498509.

Mathur, P.; Srivastava, S.; Xu, X.; and Mehta, J. L. 2020. Artificial Intelligence, Machine Learning, and Cardiovascular Disease. *Clinical Medicine Insights: Cardiology* 14: 1179546820927404.

Miotto, R.; Wang, F.; Wang, S.; Jiang, X.; and Dudley, J. T. 2018. Deep learning for healthcare: review, opportunities and challenges. *Briefings in bioinformatics* 19(6): 1236–1246.

Mozafari, M.; Farahbakhsh, R.; and Crespi, N. 2019. A BERT-based transfer learning approach for hate speech detection in online social media. In *International Conference on Complex Networks and Their Applications*, 928–940. Springer.

Ormel, J.; Von Korff, M.; Burger, H.; Scott, K.; Demyttenaere, K.; qin Huang, Y.; Posada-Villa, J.; Pierre Lepine, J.; Angermeyer, M. C.; Levinson, D.; de Girolamo, G.; Kawakami, N.; Karam, E.; Medina-Mora, M. E.; Gureje, O.; Williams, D.; Haro, J. M.; Bromet, E. J.; Alonso, J.; and Kessler, R. 2007. Mental disorders among persons with heart disease — results from World Mental Health surveys. *General Hospital Psychiatry* 29(4): 325–334. ISSN 0163-8343. doi:https://doi.org/10.1016/j.genhosppsych.2007.03.009. URL https://www.sciencedirect.com/science/article/pii/S0163834307000618.

Park, A.; Conway, M.; and Chen, A. T. 2018. Examining thematic similarity, difference, and membership in three online mental health communities from Reddit: a text mining and visualization approach. *Computers in human behavior* 78: 98–112.

Powles, J.; and Hodson, H. 2017. Google DeepMind and healthcare in an age of algorithms. *Health and technology* 7(4): 351–367.

Preoţiuc-Pietro, D.; Volkova, S.; Lampos, V.; Bachrach, Y.; and Aletras, N. 2015. Studying user income through language, behaviour and affect in social media. *PloS one* 10(9): e0138717.

Pruksachatkun, Y.; Pendse, S. R.; and Sharma, A. 2019. Moments of change: Analyzing peer-based cognitive support in online mental health forums. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–13.

Radford, A.; Wu, J.; Child, R.; Luan, D.; Amodei, D.; and Sutskever, I. 2019. Language models are unsupervised multi-task learners. *OpenAI blog* 1(8): 9.

Shatte, A. B.; Hutchinson, D. M.; and Teague, S. J. 2019. Machine learning in mental health: a scoping review of methods and applications. *Psychological medicine* 49(9): 1426–1448.

Wang, Z.; Hale, S.; Adelani, D. I.; Grabowicz, P.; Hartman, T.; Flöck, F.; and Jurgens, D. 2019. Demographic inference and representative population estimates from multilingual social media data. In *The world wide web conference*, 2056–2067.

Weathers, F. W.; Bovin, M. J.; Lee, D. J.; Sloan, D. M.; Schnurr, P. P.; Kaloupek, D. G.; Keane, T. M.; and Marx, B. P.

2018. The Clinician-Administered PTSD Scale for DSM–5 (CAPS-5): Development and initial psychometric evaluation in military veterans. *Psychological assessment* 30(3): 383.

Woolson, R. F. 2007. Wilcoxon signed-rank test. *Wiley encyclopedia of clinical trials* 1–3.

Yadav, S.; Lokala, U.; Daniulaityte, R.; Thirunarayan, K.; Lamy, F.; and Sheth, A. 2021. "When they say weed causes depression, but it's your fav antidepressant": Knowledge-aware attention framework for relationship extraction. *PLoS One* 16(3): e0248299.

Yazdavar, A. H.; Al-Olimat, H. S.; Ebrahimi, M.; Bajaj, G.; Banerjee, T.; Thirunarayan, K.; Pathak, J.; and Sheth, A. 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*, 1191–1198.