Whispers of Trauma: Leveraging Social Media for Assessing Mental Health in Victims of Childhood Sexual Abuse

Orchid Chetia Phukan[1], Rajesh Sharma[1,2], and Arun Balaji Buduru[1]

[1] IIIT-Delhi, India
{orchidp, arunb}@iiitd.ac.in
[2] University of Tartu, Estonia
rajesh.sharma@ut.ee

Abstract. ***Warning:*** *Reader discretion is recommended as our study tackles topics such as child sexual abuse, molestation, etc.*
Online Social Media (OSM) is increasingly being used by individuals with different mental health problems for finding support groups. As such extensive research has been carried out for understanding the mental health of individuals by observing their activities on OSM. However, previous studies haven't put much focus on studying mental health in victims of Childhood Sexual Abuse (CSA). CSA is a menace to society and has long-lasting effects on the mental health of the survivors. Proper care and attention towards CSA survivors facing mental health problems can drastically improve their mental health. Our work fills this gap by studying Reddit posts related to CSA. To this end, we collect and create a dataset of around 8192 CSA-related posts. For further understanding the characteristics of CSA-related posts, we performed a comparative analysis with 9198 non-CSA mental health-related posts using various natural language processing (NLP) techniques such as word-shift, word cloud, topic analysis, and emotion analysis. We found that observable differences exist between them in terms of topics being discussed and emotions. Additionally, we propose a modeling framework, ***Mental*** **F**eature-**i**nteraction **T**ransformer (***MentalFiT***) for identifying mental health problems in posts associated with CSA. It follows the notion that mental health problems and emotions are related to each other. The modeling approach involves the use of a transformer encoder to facilitate interaction among features at the intra-level, while the feature interaction block is employed to facilitate interaction among features at the inter-level. Thorough and extensive experimentation conveys the efficacy of the proposed framework on our novel curated dataset. The proposed method demonstrates impressive performance compared to models that do not incorporate emotional features. Our study opens up a new perspective toward understanding mental health problems in CSA victims and will serve as a frontier for upcoming work in this direction. Datasets and codes will be made available here[3].

[3] https://github.com/orchidchetiaphukan/CSA_ASONAM2024

Keywords: Childhood Sexual Abuse, Mental Health, Social Media, Natural Language Processing, Transformer

1 Introduction

Sexually harmful acts against a child (less than 18 years of age) are characterized as Childhood Sexual Abuse (CSA) and it spans a wide range of acts including sexual abuse, sexual interaction, rape, sexual grooming, and sexual exploitation [40]. This form of abuse often entails the perpetrator using force or making threats [5]. Between ages 7 and 13, children are most vulnerable to CSA and Perpetrators can be either family members or strangers [15].

It has been reported that survivors of CSA are often accompanied by feelings of stigma and guilt [9, 27]. CSA is often succeeded by physical and mental difficulties. Physical difficulties include physical injuries such as genital injuries [36] and sexually transmitted diseases (STDs) [25]. Individuals with a history of CSA are more likely to have chronic illnesses such as chronic pain [26] and obesity [23] in later stages of life. Physical injuries may heal, but psychological scars from the abuse can last throughout later stages of life [25]. Survivors of CSA are often likely to be haunted by the recall of the traumatic event [33]. Risk of occurrence of various mental health issues such as Post-Traumatic Stress Disorder (PTSD) [4], depression, anxiety [26], and eating disorders [56] are more in CSA survivors. CSA is also reported to be a key risk factor for the development of Borderline Personality Disorder (BPD) [1].

For the purpose of understanding mental health issues in CSA survivors, numerous prior research investigations have been conducted, but they were primarily restricted to interviews, questionnaires, surveys, and electronic health records (EHRs) with a limited amount of data [2, 55]. In recent years research considering online social media (OSM) for mental health issues is rising [7], as it has been seen that individuals are resorting to OSM platforms to find shelter for their mental health issues [6] where CSA survivors are no exception. Especially OSM such as Reddit has turned out to be a popular medium [16] for such. Reddit is a community-based platform where each community is called a "subreddit" and is based on a particular topic or issue [10]. These communities sometimes act as a discussion forum and sometimes as peer support groups for various issues related to health, mental complications, etc. Thus, this inspired us to look through various subreddits and gather posts about mental health issues associated with CSA and as far as our understanding goes, we are the pioneers in this endeavor.

We performed an extensive study of the collected posts related to CSA and propose a model architecture for the detection of mental health problems in CSA-related posts. The model entails emotional features that act as a supporting medium for the effective identification of mental health issues.

To summarize, our contributions are the following:

- Dataset of 8192 posts collected from 6040 unique users and curated from CSA-related subreddits (*r/adultsurvivors*) and different mental health subreddits (*r/ptsd*, *r/depression*, *r/depression_help*, *r/BPD*, *r/Anxiety*, *r/bipolar*, *r/schizophrenia*, and *r/mentalhealth*).

- We did a comprehensive analysis of CSA-related posts by performing a comparative study with non-CSA-related posts. We employ various natural language processing (NLP) techniques such as word-shift, word cloud, topic analysis, and emotion analysis. Our analysis points that noticeable differentiation exists between CSA and non-CSA posts in terms of the topics discussion and emotions. Mentioning flashbacks of the traumatic event and trauma-related issues are common in CSA-related issues. Emotion "fear" is observed more in CSA-related posts in contrast to non-CSA posts which are more associated with emotion "sadness". Taking these differences, we also build a model to identify CSA-related posts.
- We propose a novel framework for detecting mental health issues in CSA-related posts. We are the first, to the best of our knowledge to employ a modeling approach that takes into account the relationship between different mental health issues and emotions for identification of mental health problems in CSA-related posts. The approach comprises a transformer encoder for effective interaction of intra-level features and a feature interaction block for inter-level features. We perform extensive experiments and the proposed model shows stronger performance over models that don't leverage emotional features.

Ethical and Privacy Concerns: Although our institutional review board waived review of our work, we implemented measures to protect user privacy. No sensitive or meta information was collected or analyzed. Example posts are paraphrased to avoid traceability. The proposed predictive model processes only posts, with no user information required. Additionally, this work does not make any diagnostic claims.

2 Related Work

In this section, we first discuss existing literature related to mental health in CSA victims in Subsection 2.1 and then works that have used OSM for assessing mental health issues in Subsection 2.2.

2.1 Childhood Sexual Abuse and Mental Health

A direct relationship between CSA and severe mental health issues has been observed in previous studies [2, 42]. Studies have discovered that CSA has a strong association with serious mental health issues in women who were victims of CSA. Substance addiction and suicidal behavior were also more prevalent among CSA survivors [38]. Research focused on children exposed to CSA (both male and female, 16 years of age or younger) reported that children exposed to CSA are substantially more likely than the general population to need psychiatric treatment. Schizophrenia was the least common mental health issue among CSA survivors of all the mental health issues [47]. Men exposed to CSA are at higher risk of severe mental health issues and also suicidality [12]. Electronic health records (EHRs) were also availed for a better understanding of the relationship between CSA and serious mental health issues. It was found that patients who have been exposed to CSA are more likely to develop clinical depression, PTSD, and personality disorders [55]. Childhood exposure to sexual trauma has also been linked to the presence of bipolar disorder in later stages [14].

2.2 Online Social Media and Mental Health

Various research studies using OSM data [28, 53] for analyzing and identifying mental health issues have been conducted during the previous decade and this has proven to be an effective method [7]. These studies are divided into two groups: those that focus on a single mental health issue and those that focus on multiple mental health issues. Reddit posts were leveraged to identify depressive disorder using multiple predictive modeling approaches such as support vector machine (SVM) and multilayer perceptron (MLP) [50]. Furthermore, using a multi-modal (text, image, etc) strategy, Chiu et al. [8] investigated the identification of depressive disorder. Using tweets from self-proclaimed schizophrenia patients, Mitchell et al. [37] looked into transmitting linguistic markers for schizophrenia detection. Multiple BERT-based algorithms were also used to identify eating disorders using tweets [3].

Suicidal Intent detection is a crucial task for early detection and prevention of suicide. For identification of suicidal intent, Facebook posts were used [43]. Using RoBERTa, investigation was done for the classification of several mental health issues (depression, anxiety, bipolar disorder, Attention deficit hyperactivity disorder (ADHD), PTSD) [39]. For achieving it, Reddit posts were exploited. Previous research also looked into using a two-stage approach for detecting mental health issues, with the first stage involving binary classification into mental health or non-mental health, and the second stage involving multi-class classification of mental health issues [19, 53]. Recently COVID-19 pandemic took a toll on people's mental health which motivated researchers to look into different mental health subreddits such as *r/SuicideWatch*, *r/depression*, etc. to investigate the impact of different mental health issues on individuals during the COVID-19 pandemic's early waves [35]. Past research studies involving OSM either ignored or addressed mental health concerns in CSA victims as part of a larger study of mental health issues in the general community, with no specific emphasis placed on it. Our work takes a step in this direction by creating a dataset of posts collected from various subreddits. We carry out an extensive analysis of CSA-related posts by seeing their difference from non-CSA-related posts. Furthermore, we propose, ***MentalFiT***, for the detection of mental health issues in CSA-related posts.

3 Dataset

First, we collected posts from *r/adultsurvivors* [48] subreddit, defined as "a peer aid community for adults who experienced sexual abuse as children", which provides a platform for CSA survivors to share their experiences and seek support for their difficulties. We preferred Reddit over other OSM platforms as it is a community-focused platform and also provides a distinctive feature called "throwaway" accounts to its users. This temporary feature acts as a veil and allows the users to anonymously express their feelings or subjects that seem inexpressible [10]. As *r/adultsurvivors* is a peer support group so it may contains discussions on various mental health topics, so we filtered posts using keywords given in Table 1 related to various mental health issues such as depression, anxiety, PTSD, schizophrenia, BPD, and bipolar. We didn't collect posts from private community *r/survivorsofabuse* subreddit, a support group for abuse survivors [51]. To summarize, we only collected posts from the subreddits that are publicly accessible.

Table 1: Keywords related to different mental health problems used for collecting posts from *r/adultsurvivors*

Mental Health Problem	**Keywords**
Depression	*depression*, *depressed*, *hopelessness*, *hopeless*, *depressive*, *despondent*
Anxiety	*anxiety*, *panic attack*, *panic disorder*, *phobia*
PTSD	*ptsd*, *post-traumatic stress disorder*, *post-traumatic stess disroder*, *trauma*, *traumatic*, *traumatized*
Schizophrenia	*schizophrenia*
BPD	*bpd*
Bipolar	*bipolar*

Table 2: Data Statistics: "**# of posts**", "**# of users**" shows the count of posts collected from each subreddit and unique users

Subreddits	**# of posts**	**# of users**
adultsurvivors	6600	4567
ptsd	714	639
depression	419	400
mentalhealth	324	312
schizophrenia	45	43
depression_help	37	37
BPD	32	24
Anxiety	16	13
bipolar	5	5
Total	8192	6040

CSA victims may also seek support and discuss their mental health problems in various mental health-related subreddits. So we collected posts with CSA background from various mental health subreddits associated with a particular mental health problem such as *r/depression*, *r/Anxiety*, *r/ptsd*, and *r/depression_help* [39], *BPD*, *schizophrenia*, *bipolar* [28] using keywords *childhood sexual abuse* and *csa*. We also collected posts *r/mentalhealth* (most popular health-related subreddit) [18] using the same keywords used for collecting posts in *r/adultsurviors* together with keywords such as *childhood sexual abuse* and *csa* to make sure they are related to CSA. We collected posts posted by normal accounts as well as throwaway accounts. Posts from the subreddits are collected using PushShift API. Posts from the above-mentioned subreddits are from the date these subreddits started till January 2022. These posts are subjected to a manual inspection to ensure that they are CSA-related. Table 2 shows information about posts collected through various subreddits.We dropped posts tagged as "[deleted]" because these posts are either deleted or posted by users who have deleted their accounts and also posts tagged as "[removed]" are dropped as these posts are removed by the moderator of the subreddit or spam filter.

4 Analysis of CSA-related Posts

Mental health problems in users might not be necessarily due to CSA, so for a better understanding of the unique characteristics of posts associated with CSA and what sets them apart, we do a comparative analysis with posts not related to CSA. For example, comparing the CSA-related and non-CSA-related posts shown in Table 3, it is evident that they are pretty similar though small differences may exist. So first, we use an existing dataset [28] annotated for various mental health issues for collecting posts unrelated to CSA. The dataset contains posts from various mental health subreddits, so it may have posts with CSA background. Posts that contains words closely related with CSA such as *childhood sexual abuse* and *csa* are removed and finally 9198 posts are extracted from the dataset. These posts are again manually inspected to ensure they are not CSA-related. We use various NLP approaches such as word-shift, wordcloud, topic analysis, and emotion analysis, which are followed by previous studies [13, 41].

Table 3: Example of posts with and without CSA background

CSA-related	non-CSA-related
I've sobbed so much in the last 24 hours about my csa and its ramifications that I'm not sure I have any more tears. Depression has returned.	*I let go of a buddy who was causing me pain. It's more lonely than before. My best buddy steadily moving away from me and I've been really dreadful.*

At first, we utilize proportion shift, a type of word-shift [17] graph that ranks words based on scores generated from the difference between a word's relative frequency in the first textual content and its relative frequency in the second textual content.

As visualized in Figure 1, words that describe CSA ("abuse", "sexual", and "child") are prominent in posts with CSA background and this helps to paint a clear picture of how CSA could be the core cause of mental health issues in CSA survivors. In contrast to non-CSA posts, words that represent memory or flashbacks of the traumatic event ("remember", "happened", "memories") are prominent in posts with CSA background. From these words, it can be deduced that memories or flashbacks of the traumatic event can play a vital role in the development of mental health issue such as PTSD ("trauma") in CSA survivors [44]. In contrast to posts without CSA background, words that demonstrate familial relationships ("dad", "brother", and "mom") stand out in posts with CSA background which may point out the likelihood of perpetrators being family members [15]. For example, phrases such as "..cried..over..**csa**..", "..**memory**..being..raped..", "..**remembered**..stuff.." appeared often in the corpus with CSA background.

Words associated with the workplace ("work" and "job") are more prominent in posts without CSA background than in posts with CSA background, indicating that the development of mental health difficulties may be linked to the workplace. Words "anxiety" and "depression" are more common in non-CSA posts, indicating that these mental health concerns are more prevalent. For example, non-CSA background, "..**depression** comes back..", "..**depression** suicidal thoughts..f**king **feeling**..", "need..**friend**..".

For both kinds of posts, that is, CSA and non-CSA, we also intend to look at unique unigrams and bigrams. To score unigrams and bigrams, TF-IDF (Term frequency-Inverse document frequency) is used and the results are displayed as a wordcloud. Figure 2 presents the top ten unique unigrams and Figure 3 presents the top ten unique bigrams for CSA and non-CSA posts. These results converges with those shown in Figure 1 but the method used to score the unigrams and bigrams is different than the method used in word-shift. Additionally, bigram such as "self harm" related to suicidal intent is more prevalent in posts with CSA background while "self-esteem" appears in posts without CSA background which may direct towards mental health problems related to one's self-worth.

Topic analysis is also done for discovering various topics in CSA and non-CSA posts using BERTopic [22] and words inside each topic are scored using c-TF-IDF (class-based TF-IDF). Two topics are identified in CSA posts based on coherence score. Topic 1 is associated with CSA, such as cause ("abuse") and its after-effect ("trauma"). Topic 2 is concerned with clinical aspects ("doctor", "appointment"). Table 4 displays the topics and top five N-grams for CSA posts. In non-CSA posts thirty topics are discovered based on coherence score, but we only provide the top six topics from the thirty topics obtained and their top five N-grams in Table 5. Various topics are discovered including mental health concerns relating to the workplace ("anxiety", "depression", "work"), eating disorders ("anorexia"), and substance use ("cbd", "cbd oil", "weed yesterday").

Lastly, we pick an off-the-shelf fine-tuned DistilRoBERTa [24] model from *Hugginface* for detecting emotions to examine if the emotions being present in CSA posts are different from non-CSA posts as used by [30, 46] in a similar application. This model labels the posts to the most probable emotion among the emotions, namely, anger, sadness, fear, disgust, neutral, surprise, and joy. Figure 4 shows the emotions for CSA and non-CSA posts. Emotion "fear" is associated more with CSA posts than with non-CSA posts as strong correlation is observed with "fear" and trauma-related issues [21] and the presence of trauma-related issues is more in CSA posts "sadness" emotion is reported more in non-CSA posts as "sadness" is associated with depression and it is one of the most common mental health issue.

4.1 Identification of CSA-related posts

As seen above, subtle differences exist in CSA and non-CSA related posts. So we build a modeling approach to identify posts associated with CSA (Figure 5a). We leverage pre-trained models (PTMs) such as BERT [11], RoBERTa [32], and GPT-2 [45] as backbone architecture to retrieve text features. These models were

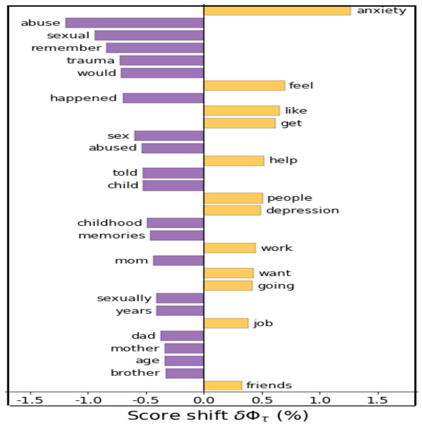
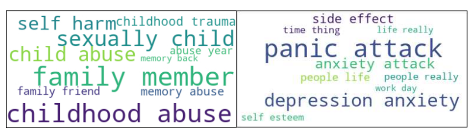


Fig. 1: Word shift for CSA posts *(left)* and non-CSA posts *(right)*

(a) CSA (b) non-CSA

Fig. 2: Wordcloud for top unigrams of CSA and non-CSA posts



(a) CSA (b) non-CSA

Fig. 3: Wordcloud for top bigrams of CSA and non-CSA posts

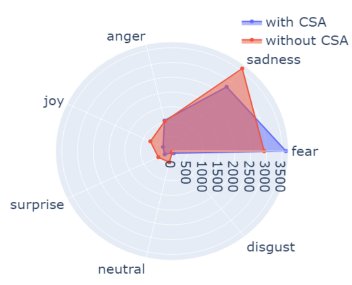


Fig. 4: Emotions of CSA and non-CSA posts

Table 4: Top topics and its top five N-grams for CSA posts

Topic 1		Topic 2	
N-gram	**Score**	**N-gram**	**Score**
abuse	0.0131	doctor	0.0924
really	0.0126	exam	0.0545
family	0.0124	appointment	0.0256
sex	0.0111	pain	0.0156
parent	0.0103	advice	0.0119

Table 5: Top topics and its top five N-grams for non-CSA posts

Topic 1		Topic 2	
N-gram	**Score**	**N-gram**	**Score**
anxiety	0.0096	bed	0.0319
depression	0.0067	anxiety	0.0147
work	0.0067	week sleep	0.0078
school	0.0062	trouble	0.0069
problem	0.0047	time sleep	0.0069
Topic 3		**Topic 4**	
N-gram	**Score**	**N-gram**	**Score**
calorie	0.0207	nausea	0.0502
weight loss	0.0144	symptom	0.0159
fat	0.0116	vomit	0.0150
exercise	0.0096	lack appetite	0.0083
anorexia	0.0094	nausea anxiety	0.0079
Topic 5		**Topic 6**	
N-gram	**Score**	**N-gram**	**Score**
ugly	0.0185	weed	0.0621
insecure	0.0148	cbd	0.0571
appearance	0.0140	cbd oil	0.0253
body dysmorphia	0.0103	anxiety depression	0.0130
self image	0.0097	weed yesterday	0.0111

Table 6: Evaluation Results of Models for Identification of Posts associated with CSA; E(BERT), E(RoBERTa), E(GPT-2) represent the models with input features from BERT, RoBERTa, and GPT-2 respectively

Model	Accuracy	F1-score
E(BERT)	0.9226	0.9224
E(RoBERTa)	0.9223	0.9220
E(GPT-2)	0.9281	0.9279
CSAFiTv1	0.9270	0.9267
CSAFiTv2	0.9252	0.9250
CSAFiTv3	**0.9312**	**0.9310**

trained on large amounts of data in an unsupervised fashion to learn general-purpose knowledge and yield richer representations. Such PTMs were exploitated by previous studies for similar applications [3, 52]. We use BERT, RoBERTa, and GPT-2 base versions available in *Hugginface*. The final hidden states are extracted from these PTMs and converted to a vector of 768-dimension through average pooling. Extracted features are fed to convolution block which comprises a 1D-CNN followed by a maxpooling layer for further retaining important features. This is followed by a transformer encoder that comprises sub-blocks: Multi-Head Attention and Feed-forward layers as proposed by [54]. Its output is flattened and then passed through a classifier block which basically consists of multiple Fully-Connected (FCN) layers. The last layer of the classifier block is succeeded by *Softmax* activation function which outputs probabilities for each class i.e. either CSA-related or not. We use *Cross-entropy* as the loss function and *Adam* as the optimizer.

Data Preprocessing: We convert the posts to lowercase followed by removing punctuation marks and stopwords.

Training Details: Total posts sum upto to 17390 (8192 related to CSA and 9198 not related to CSA). We randomly split these posts in the ratio of 80:20 to training and test set. Again, 10% of the training set is kept for validation. We train the models for 20 epoch each with a learning rate of 1e-3 and batch size of 16. We use Accuracy and F1-Score (Macro) as evaluation metrics.

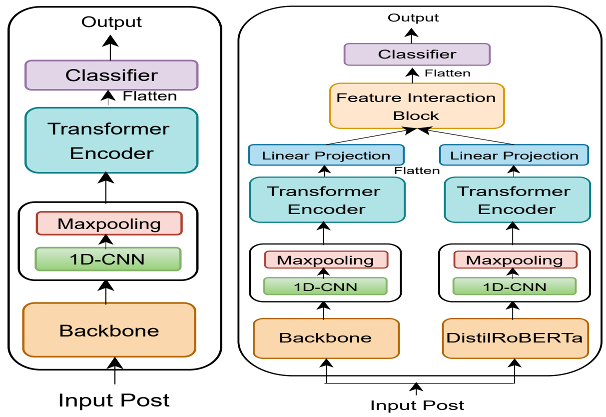
Results: The results of the models are shown in Table 6. The model with GPT-2 features performed the best across models that uses text features from different PTMs. We also implemented models that leverage both PTM and emotion features from DistilRoBERTa as in Section 5 while following the same modeling approach (*CSAFiTv1*, *CSAFiTv2*, *CSAFiTv3* uses BERT, RoBERTa, GPT-2 embeddings as input features respectively). We can see from the results, adding emotion does lead to improvement in performance but with minimal difference. *CSAFiTv3* reports the best accuracy (93.12%) and F1-Score (93.10%) among all the models.

5 Detection of Mental Health Problems in CSA Posts

5.1 Model Architecture

We propose a model architecture ***Mental*** **F**eature-**i**nteraction **T**ransformer (***MentalFiT***) for detecting mental health problems in CSA-related posts (Figure 5b). Its three variants are ***MentalFiTv1***, ***MentalFiTv2***, and ***MentalFiTv3***. These model variants comprise of a transformer encoder with a multi-head attention sub-block that learns the intra-level as used by prior studies in similar applications related to mental health problems detectoin [34, 49] and a feature interaction block (FIB) that learns the inter-level interaction of input features. We consider two types of input features. First, for general-purpose representations, we extract text features from PTMs BERT, RoBERTa, and GPT-2. The difference in the proposed model variants arises from the PTMs used for retrieving features. ***MentalFiTv1***, ***MentalFiTv2***, ***MentalFiTv3*** uses BERT, RoBERTa, and GPT-2 respectively.

Additionally, as emotions and mental health problems are intertwined with each other and different mental health conditions are associated with different



(a) (b)

Fig. 5: Model Architecture; Backbone here represents different PTMs (BERT, RoBERTa, GPT-2); In (*Left*), model architecture without Emotion Features; In (*Right*), model architecture with Emotion Features leveraged from emotion fine-tuned DistilRoBERTa

emotions [20], we leverage fine-tuned DistilRoBERTa [24] for emotional text features as the second type of input. Both from the PTMs and DistilRoBERTa last hidden representations are extracted and converted to 768-dimension vectors through average pooling. These features are individually passed through a convolution block followed by a transformer encoder similarly as in Section 4.1. The output from the encoders is flattened and is linearly projected to 120-dimension. It is then fed to the FIB which performs the outer product of the incoming inputs leading to effective fusion [29] of the features (general purpose representations from PTMs and emotion-based features from fine-tuned emotoin recognition model). This will result in a matrix of the shape of 120×120. The linear projection is performed as output from the transformer encoders are of dimension 12288 and the outer product will result in a matrix of shape 12288×12288 and this calls for a model of billion parameters. This scenario is out of bounds due to computational requirements. Output from FIB is flattened and fed to a classifier that comprises multiple FCNs. The output layer is passed through *Softmax* to access the probabilities i.e different mental health problems. We use *Focal Loss* [31] (Equation 1) as the loss function and *Adam* as the optimization function.

$$FL(p_t) = -\alpha(1-p_t)^{\gamma}\ln(p_t) \tag{1}$$

where, $FL(p_t)$ represents the Focal Loss, p_t denotes the predicted probability of the true class; α is the balancing factor that can be assigned to each class to regulate how much weight is given to each class; γ is focusing parameter that controls the focusing level of the loss; $-\ln(p_t)$ represents cross-entropy loss;

$(1-p_t)^\gamma$ stands for focusing factor that emphasizes more on hard-to-classify examples and less on the easy ones.
Data Preprocessing: As followed in Section 4.1.
Training Details: We collect posts from subreddits *r/ptsd*, *r/depression*, *r/depression_ help*, *r/BPD*, *r/Anxiety*, *r/bipolar*, and *r/schizophrenia*, each associated with a specific mental health problem. However, for *r/adultsurvivors* and *r/mentalhealth*, there are no such labels available. The total number of posts from these subreddits sums up to 1268. We label the posts from *r/BPD*, *r/Anxiety*, *r/bipolar*, and *r/schizophrenia* as "others" due to low number of posts related to CSA while the other two classes being PTSD (*r/ptsd*) and Depression (*r/depression*, *r/depression_ help*) framing the problem as a three-class classification. As this is a case of extreme class imbalance, so focal loss is preferred. We split the posts into training and test sets with 80:20 ratio and 10% from the training set is extracted as validation set. We train the models for 20 epochs with learning rate and batch size set as 1e-3 and 16 respectively. Evaluation metrics are used similar to Section 6. We set the hyperparameters based on *Gridsearch*.

5.2 Results:

Evaluation of models for detection of mental health problems in CSA-related posts is shown in Table 7. We also implement models used in Section 6 that employs transformer encoder preceded by features from PTMs (BERT, RoBERTa, GPT-2). Among these, E(GPT-2) model performs the best achieving accuracy of 76.77% and 56.11% F1-Score. The proposed models employing emotional features improve over other models leveraging only PTM features by a margin of 3-5% in terms of accuracy and F1-Score. This proves the efficacy of the approach and that emotional features act as complementary information for mental health issues detection. Additionally, we can visualize improvement in F1-Score due to the application of focal loss that helps in mitigating the problem of class imbalance to a certain extent. Among the different proposed variants, *MentalFiTv2* retained the top position in accuracy with 80.31% and *MentalFiTv3* in F1-Score with 64.35%.

Table 7: Evaluation Results of models for detecting mental health problems in CSA-related posts; E(BERT), E(RoBERTa), E(GPT-2) represent the models with input features from BERT, RoBERTa, and GPT-2 respectively

Model	Accuracy	F1-score
E(BERT)	0.7559	0.5516
E(RoBERTa)	0.7441	0.5452
E(GPT-2)	0.7677	0.5611
MentalFiTv1	0.7834	0.5981
MentalFiTv2	**0.8031**	0.5865
MentalFiTv3	0.7913	**0.6435**

6 Conclusion

Our work contributes by exploiting OSM for the study of mental health issues in CSA survivors as previous studies have not paid particular attention towards them. We created a dataset by collecting mental health issues related posts related to CSA not only from CSA-related subreddit but also from multiple mental-health-related subreddits. Our research towards gaining an understanding of uniqueness of CSA-related posts led to the discovery of existence of minor variations between CSA and non-CSA posts . Furthermore, we introduce a modeling framework, *MentalFiT*, designed to detect mental health issues in posts linked to CSA. This model operates under the premise that there exists a connection between mental health problems and emotions. The modeling strategy employs a transformer encoder to encourage interaction among features within the same set of features, while the feature interaction block promotes interaction across different sets of features. Extensive and thorough evaluation supports the usefulness of the proposed framework, which performs admirably when compared to models that do not incorporate emotional characteristics. In our future work, we would like to include additional subreddits related to mental issues in our analysis. One of the reasons for including more subreddits is to include additional mental health issues in our analysis.

References

1. de Aquino Ferreira, L.F., Pereira, F.H.Q., Benevides, A.M.L.N., Melo, M.C.A.: Borderline personality disorder and sexual abuse: a systematic review. Psychiatry research **262**, 70–77 (2018)
2. Banyard, V.L., Williams, L.M., Siegel, J.A.: The long-term mental health consequences of child sexual abuse: An exploratory study of the impact of multiple traumas in a sample of women. Journal of traumatic stress **14**(4), 697–715 (2001)
3. Benítez-Andrades, J.A., Alija-Pérez, J.M., García-Rodríguez, I., Benavides, C., Alaiz-Moretón, H., Vargas, R.P., García-Ordás, M.T.: Bert model-based approach for detecting categories of tweets in the field of eating disorders (ed). In: 2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS). pp. 586–590. IEEE (2021)
4. Briere, J., Runtz, M.: Post sexual abuse trauma: Data and implications for clinical practice. Journal of interpersonal violence **2**(4), 367–379 (1987)
5. Browne, A., Finkelhor, D.: Impact of child sexual abuse: A review of the research. Psychological bulletin **99**(1), 66 (1986)
6. Bucci, S., Schwannauer, M., Berry, N.: The digital revolution and its impact on mental health care. Psychology and Psychotherapy: Theory, Research and Practice **92**(2), 277–297 (2019)
7. Chancellor, S., De Choudhury, M.: Methods in predictive techniques for mental health status on social media: a critical review. NPJ digital medicine **3**(1), 1–11 (2020)
8. Chiu, C.Y., Lane, H.Y., Koh, J.L., Chen, A.L.: Multimodal depression detection on instagram considering time interval of posts. Journal of Intelligent Information Systems **56**(1), 25–47 (2021)
9. Coffey, P., Leitenberg, H., Henning, K., Turner, T., Bennett, R.T.: Mediators of the long-term impact of child sexual abuse: Perceived stigma, betrayal, powerlessness, and self-blame. Child Abuse & Neglect **20**(5), 447–455 (1996)
10. De Choudhury, M., De, S.: Mental health discourse on reddit: Self-disclosure, social support, and anonymity. In: Eighth international AAAI conference on weblogs and social media (2014)

11. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)
12. Easton, S.D., Renner, L.M., O'Leary, P.: Suicide attempts among men with histories of child sexual abuse: Examining abuse severity, mental health, and masculine norms. Child Abuse & Neglect **37**(6), 380–387 (2013)
13. Elmas, T., Overdorf, R., Aberer, K.: Characterizing retweet bots: The case of black market accounts. In: Proceedings of the International AAAI Conference on Web and Social Media. vol. 16, pp. 171–182 (2022)
14. Fijtman, A., Clausen, A., Kauer-Sant'Anna, M., Workgroup, V.M.A.M., Morey, R.: Trauma history in veterans with bipolar disorder and its impact on suicidality. Journal of psychiatric research **157**, 119–126 (2023)
15. Finkelhor, D.: Current information on the scope and nature of child sexual abuse. Future Child. **4**(2), 31–53 (1994)
16. Fraga, B.S., da Silva, A.P.C., Murai, F.: Online social networks in health care: a study of mental disorders on reddit. In: 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI). pp. 568–573. IEEE (2018)
17. Gallagher, R.J., Frank, M.R., Mitchell, L., Schwartz, A.J., Reagan, A.J., Danforth, C.M., Dodds, P.S.: Generalized word shift graphs: a method for visualizing and explaining pairwise comparisons between texts. EPJ Data Science **10**(1), 4 (Jan 2021). https://doi.org/10.1140/epjds/s13688-021-00260-3, `https://doi.org/10.1140/epjds/s13688-021-00260-3`
18. Gaur, M., Kursuncu, U., Alambo, A., Sheth, A., Daniulaityte, R., Thirunarayan, K., Pathak, J.: " 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. pp. 753–762 (2018)
19. Gkotsis, G., Oellrich, A., Velupillai, S., Liakata, M., Hubbard, T.J., Dobson, R.J., Dutta, R.: Characterisation of mental health conditions in social media using informed deep learning. Scientific reports **7**(1), 1–11 (2017)
20. Gloria, C.T., Steinhardt, M.A.: Relationships among positive emotions, coping, resilience and mental health. Stress and health **32**(2), 145–156 (2016)
21. Gonzalez, P., Martinez, K.G.: The role of stress and fear on the development of psychopathology. The Psychiatric Clinics of North America **37**(4), 535 (2014)
22. Grootendorst, M.: Bertopic: Neural topic modeling with a class-based tf-idf procedure. arXiv preprint arXiv:2203.05794 (2022)
23. Gustafson, T., Sarwer, D.: Childhood sexual abuse and obesity. Obesity reviews **5**(3), 129–135 (2004)
24. Hartmann, J.: Emotion english distilroberta-base. `https://huggingface.co/j-hartmann/emotion-english-distilroberta-base/` (2022)
25. Johnson, C.F.: Child sexual abuse. The Lancet **364**(9432), 462–470 (2004)
26. Kamiya, Y., Timonen, V., Kenny, R.A.: The impact of childhood sexual abuse on the mental and physical health, and healthcare utilization of older adults. International psychogeriatrics **28**(3), 415–422 (2016)
27. Kennedy, A.C., Prock, K.A.: "i still feel like i am not normal": A review of the role of stigma and stigmatization among female survivors of child sexual abuse, sexual assault, and intimate partner violence. Trauma, Violence, & Abuse **19**(5), 512–527 (2018)
28. Kim, J., Lee, J., Park, E., Han, J.: A deep learning model for detecting mental illness from user content on social media. Scientific Reports **10**(1), 11846 (Jul

2020). https://doi.org/10.1038/s41598-020-68764-y, https://doi.org/10.1038/s41598-020-68764-y
29. Kumar, G.K., Nanadakumar, K.: Hate-clipper: Multimodal hateful meme classification based on cross-modal interaction of clip features. arXiv preprint arXiv:2210.05916 (2022)
30. Li, R., Sun, S., Elhoseiny, M., Torr, P.: Oxfordtvg-hic: Can machine make humorous captions from images? arXiv preprint arXiv:2307.11636 (2023)
31. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE international conference on computer vision. pp. 2980–2988 (2017)
32. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V.: Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692 (2019)
33. Loftus, E.F., Polonsky, S., Fullilove, M.T.: Memories of childhood sexual abuse: Remembering and repressing. Psychology of women quarterly **18**(1), 67–84 (1994)
34. Lokala, U., Srivastava, A., Dastidar, T.G., Chakraborty, T., Akhtar, M.S., Panahiazar, M., Sheth, A.: A computational approach to understand mental health from reddit: knowledge-aware multitask learning framework. In: Proceedings of the International AAAI Conference on Web and Social Media. vol. 16, pp. 640–650 (2022)
35. Low, D.M., Rumker, L., Talkar, T., Torous, J., Cecchi, G., Ghosh, S.S.: 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 (2020)
36. McCann, J., Voris, J., Simon, M.: Genital injuries resulting from sexual abuse: a longitudinal study. Pediatrics **89**(2), 307–317 (1992)
37. Mitchell, M., Hollingshead, K., Coppersmith, G.: Quantifying the language of schizophrenia in social media. In: Proceedings of the 2nd workshop on Computational linguistics and clinical psychology: From linguistic signal to clinical reality. pp. 11–20 (2015)
38. Mullen, P.E., Martin, J.L., Anderson, J.C., Romans, S.E., Herbison, G.P.: Childhood sexual abuse and mental health in adult life. The British Journal of Psychiatry **163**(6), 721–732 (1993)
39. Murarka, A., Radhakrishnan, B., Ravichandran, S.: Detection and classification of mental illnesses on social media using roberta. arXiv preprint arXiv:2011.11226 (2020)
40. Murray, L.K., Nguyen, A., Cohen, J.A.: Child sexual abuse. Child and Adolescent Psychiatric Clinics **23**(2), 321–337 (2014)
41. Neha, K., Agrawal, V., Kumar, V., Mohan, T., Chopra, A., Buduru, A.B., Sharma, R., Kumaraguru, P.: A tale of two sides: Study of protesters and counter-protesters on# citizenshipamendmentact campaign on twitter. In: 14th ACM Web Science Conference 2022. pp. 279–289 (2022)
42. Nelson, E.C., Heath, A.C., Madden, P.A., Cooper, M.L., Dinwiddie, S.H., Bucholz, K.K., Glowinski, A., McLaughlin, T., Dunne, M.P., Statham, D.J., et al.: Association between self-reported childhood sexual abuse and adverse psychosocial outcomes: results from a twin study. Archives of general psychiatry **59**(2), 139–145 (2002)
43. Ophir, Y., Tikochinski, R., Asterhan, C.S., Sisso, I., Reichart, R.: Deep neural networks detect suicide risk from textual facebook posts. Scientific reports **10**(1), 1–10 (2020)

44. Priebe, K., Kleindienst, N., Zimmer, J., Koudela, S., Ebner-Priemer, U., Bohus, M.: Frequency of intrusions and flashbacks in patients with posttraumatic stress disorder related to childhood sexual abuse: an electronic diary study. Psychological assessment **25**(4), 1370 (2013)
45. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I., et al.: Language models are unsupervised multitask learners. OpenAI blog **1**(8), 9 (2019)
46. Rozado, D., Hughes, R., Halberstadt, J.: Longitudinal analysis of sentiment and emotion in news media headlines using automated labelling with transformer language models. Plos one **17**(10), e0276367 (2022)
47. Spataro, J., Mullen, P.E., Burgess, P.M., Wells, D.L., Moss, S.A.: Impact of child sexual abuse on mental health: prospective study in males and females. The British Journal of Psychiatry **184**(5), 416–421 (2004)
48. Stanton, A.M., Meston, C.M., Boyd, R.L.: Sexual self-schemas in the real world: Investigating the ecological validity of language-based markers of childhood sexual abuse. Cyberpsychol. Behav. Soc. Netw. **20**(6), 382–388 (Jun 2017)
49. Sun, H., Liu, J., Chai, S., Qiu, Z., Lin, L., Huang, X., Chen, Y.: Multi-modal adaptive fusion transformer network for the estimation of depression level. Sensors **21**(14), 4764 (2021)
50. Tadesse, M.M., Lin, H., Xu, B., Yang, L.: Detection of depression-related posts in reddit social media forum. IEEE Access **7**, 44883–44893 (2019)
51. TeBlunthuis, N., Hill, B.M.: Identifying competition and mutualism between online groups. In: Proceedings of the International AAAI Conference on Web and Social Media. vol. 16, pp. 993–1004 (2022)
52. Triantafyllopoulos, I., Paraskevopoulos, G., Potamianos, A.: Depression detection in social media posts using affective and social norm features. arXiv preprint arXiv:2303.14279 (2023)
53. Vajre, V., Naylor, M., Kamath, U., Shehu, A.: Psychbert: A mental health language model for social media mental health behavioral analysis. In: 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp. 1077–1082. IEEE (2021)
54. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)
55. Werbeloff, N., Hilge Thygesen, J., Hayes, J.F., Viding, E.M., Johnson, S., Osborn, D.P.: Childhood sexual abuse in patients with severe mental illness: Demographic, clinical and functional correlates. Acta Psychiatrica Scandinavica **143**(6), 495–502 (2021)
56. Wonderlich, S.A., Brewerton, T.D., Jocic, Z., Dansky, B.S., Abbott, D.W.: Relationship of childhood sexual abuse and eating disorders. Journal of the American Academy of Child & Adolescent Psychiatry **36**(8), 1107–1115 (1997)