Event Detection for Suicide Understanding

Anonymous ACL submission

Abstract

Suicide is a serious problem in every society. Understanding life events of a potential patient is essential for successful suicide-risk assessment and prevention. In this work, we focus on the Event Detection (ED) task to identify event trigger words of suicide-related events in public posts of discussion forums. In particular, we introduce a new dataset for ED (called SuicideED) that features seven suicidal event types to comprehensively capture suicide actions and ideation, and general risk and protective factors. Our experiments with current state-of-the-art ED systems suggest that there is still room for improvement of ED models in this domain. We will publicly release SuicideED to support future research in this important area.

1 Introduction

Suicide is a serious and growing problem in our society[1]. The most common procedure for suicide risk assessment is for clinicians to set up clinical interviews with potential patients that will provide rating scales based on a list of preset questions (Ross et al., 2012). However, interviews and similar activities require the willingness of potential patients to participate. Given the associated mental states, such participation can be challenging to obtain for patients with high suicidal risks.

In the meantime, people are increasingly spending more of their time on social networks, sharing inner thoughts and daily activities. This collection of social posts might draw a comprehensive picture of the patient's life that can be used to support the diagnosis of suicidal conditions. In fact, moderators of some social networks (e.g., Reddit, Reachout) use social posts to monitor suicide and apply immediate intervention if necessary. To assist with the processing of the large amount of posts, there have been a few methods and tools for automatically analyzing online posts to detect suicidal intent (Ji et al., 2018; Shing et al., 2018; Coppersmith et al., 2015; Milne et al., 2016). However, these studies mainly focus on assessing the patients' susceptibility to suicide and fail to consider contributory life events that cause/lead to such conditions. In this paper, we argue that recognizing suicide-related events is also critical to suicide understanding, identification and prevention, and natural language processing (NLP) methods are necessary to support automatic identification of such events from the vast and growing number of social media posts.

[1] https://www.nimh.nih.gov/health/statistics/suicide

This work aims to advance the ultimate goal of creating NLP methods for suicide understanding by exploring the novel task of Event Detection (ED) for suicide-related events. ED is an important task in Information Extraction (IE) that aims to identify event trigger words/mentions in the text (Ahn, 2006; Ji and Grishman, 2008). Adapted to our interest in suicide-related events, in the following sentences, an ED system should be able to recognize "*date*" and "*have*" as trigger words for deteriorated personal relationship events (i.e., risk factors); "*wanna*", "*have*", and "*desirable*" as triggers for protective factor events, and "*depression*" as a trigger for a health-related risk factor event:

I don't date anyone and never will. It s a reason why I have no friends and never will. I wanna be funny and have a personality and be desirable but I 'm not that. I know its depression that causes it.

The vast majority of advanced methods for ED are based on training deep neural networks on large labeled corpora (Chen et al., 2015). As such, to facilitate ED in suicide prevention research, a key requirement is to have a benchmark dataset to standardize the development and evaluation of ED models. Unfortunately, a large amount of existing suicide-related datasets are protected due to sensitive privacy concerns and, thus, fail to support the larger research community (Coppersmith et al., 2015; Vioules et al., 2018; Bhat and Goldman-Mellor, 2017). Moreover, these existing datasets are created to detect potential suicidal attempts based on text classification (Vioules et al., 2018; Bhat and Goldman-Mellor, 2017; Shing et al., 2018), which does not provide event trigger annotations of suicide-related events for ED.

To overcome such challenges, this paper introduces SuicideED, a new dataset for suicidal event detection that is manually annotated for seven distinct event types to comprehensively characterize suicide-related events regarding actual actions, thoughts, and risk and protective factors. To enable data sharing, our dataset is based on pubic posts from Reddit where personal information is not presented to avoid privacy issues. The SuicideED dataset is challenging as it involves informal texts, and require event factuality and affected entity reasoning. Our experiments show that the performance of current state-of-the-art ED models on SuicideED lags behind their performance on other general-purpose ED datasets, thus calling for more research effort for suicide-specific ED. To facilitate future research in this area, SuicideED will be released publicly for the research community.

2 Annotation

The documents for SuicideED are collected out of publicly available posts from `reddit.com`. In particular, we focus on three subreddits (subgroups) that contain a high percentage of suicide-related posts: `r/SuicideWatch`, `r/depression`, and `r/mentalhealth`. Each original post is considered as a separate document and only posts with more than 50 words are kept to increase the probability of an event being present.

Event Types: An important, previously unexplored, question is *what* constitutes relevant events that can provide useful insights for clinicians to better understand and recognize suicide-related incidents. Accordingly, we consult specialized literature related to suicidal-behavior identification and treatment (Gutierrez, 2006; de Ruiter and Nicholls, 2011; O'Connor et al., 2013) to define the event categories for our dataset. As such, we design the event types to be exclusive to avoid type overlapping, and sufficiently comprehensive to cover relevant/impactful suicide-related events in the data. Eventually, we select the following seven event types that capture suicide-related actions, thoughts, and risk/protective factors.

The first two event types are concerned with statements to indicate suicidal attempts or intentions. In particular, the **ACTION** event type is dedicated to the direct expressions for actual suicidal attempts/actions, e.g., "*I've started* ***cutting*** *myself again*"[2]. In contrast, the second event type, **IDEATION**, represents suicidal inner thoughts, feelings, or desires, where no real action present, e.g., "*I'm going to* ***kill*** *myself soon*". These two types directly integrate factuality differentiation into the event types to better address the uniqueness of the data where hypothetical events are prevalent and understanding the factuality of events is critical to suicide intervention and prevention.

The second group of event types focuses on external events that increase a subject's susceptibility to suicidal behaviours, i.e., risk factors (RF) (Gutierrez, 2006). Given the diverse nature of RF, four event types are proposed. **RF-LIFE** events include mentions of a death of a close/loved entity, e.g., "*My dog just* ***died***". **RF-RELATIONSHIP** concerns events related to social isolation, family breakdowns, or any mention of deteriorated interpersonal relationship, e.g., "*My dad* ***kicked*** *me out of the house*". Events for **RF-HEALTH** cover mentions of physical diseases, mental illness, and behaviors that directly affect the subject's health, e.g., "*I feel* ***depressed***". Finally, **RF-OTHER** incorporates every other RF event that cannot be assigned to life, relationship, or health issues but still qualify as RF, including financial issues, chronic abuse, and general quality-of-life problems.

The final type, **PROTECTIVE**, captures events that drive an individual towards a better mental-health state, involving a broad range of positive activities, such as receiving effective medication or being motivated by social connections, e.g., "*The medication seems to be* ***helping***". A detailed description and representative examples for each event type are presented in Appendix C.

Annotation: Given the event types, an annotation job posting is created on `upwork.com` and seven freelance annotators with previous experience in mental health and psychology, such as physicians and psychology graduates, are recruited. They are provided with a comprehensive guideline document with thorough annotation instructions and numerous detailed examples for training. The annotators are instructed to select a single word for each event trigger (i.e., the most im-

[2]In our examples, event trigger words are highlighted.

	Train	Dev	Test
#Event triggers	33,055	1,925	1,998
#Documents	2,214	130	109
#Sentences	20,677	1,178	1,176
#Words	378,435	20,301	21,541

Table 1: Data statistics for SuicideED.

Label	Count
RF-OTHER	15,343
PROTECTIVE	7,389
IDEATION	6,645
RF-RELATION	3,890
RF-HEALTH	2,408
ACTION	1,084
RF-LIFE	219

Table 2: Label distribution of SuicideED.

portant) that clearly evokes the event, following the practices of prior ED work (Nguyen and Grishman, 2015). Overall, we annotate 2,300 documents for the seven event types from which the proportions of documents taken from the subreddits `r/SuicideWatch`, `r/depression`, and `r/mentalhealth` are 50%, 30%, and 20%, respectively. 20% of the documents are selected for co-annotation, leading to a Fleiss' Kappa score of 0.8 (i.e., close to the almost perfect agreement range of $[0.81 - 1.0]$). The remaining 80% of documents are distributed to the annotators for individual annotation. To facilitate future research, we divide SuicideED into three different portions for training, test, and development data. Table 1 presents some statistics for different data portions while Table 2 shows the event type distribution.

3 Dataset Challenges

Compared to existing ED datasets, e.g., ACE-05 (Walker et al., 2006), MAVEN (Wang et al., 2020), and CySecED (Trong et al., 2020), our dataset SuicideED features at least three unique challenges for ED models. First, as its documents are obtained from Reddit posts, SuicideED involves texts where informal words (e.g., "*wanna*", "*gonna*") are prevalent, sentences might not follow well-structured grammar rules, and first-person point of view is the main writing style. This is in contrast to existing ED datasets where documents are often retrieved from news or reports with formal and well-structured texts. Second, in addition to the relevance to suicide, the event types in SuicideED sometimes require models to simultaneously consider event factuality to accurately determine the types. This is clear for differentiating **ACTION** and **IDEATION** where the key distinction concerns event factuality. Another example involves the potential confusions between **PROTECTIVE** and **RF** where different event factuality might lead to different event types for the similar expressions. For instance, in the sentence "*I **have** a lot of friends*", the event trigger "*have*", belongs to the **PROTECTIVE** type given that it reveals a positive environment for the subject. On the contrary, in the sentence "*I **had** a lot of friends*", the trigger "*had*" should be considered as a **RF-RELATIONSHIP** type as it might instead imply current deteriorated social connections. Last but not least, the event type determination in SuicideED also necessitates appropriate identification of the entity that should be considered for the effect of an event. For instance, in the sentence "*My sister **killed** herself.*", the trigger word "*killed*" should have the **ACTION** type if the entity of consideration is "*sister*". However, if we consider the event from the point of view of the poster/speaker, "*killed*" should be a **RF-LIFE** event. In SuicideED, the annotators are instructed to take the first person point of view (i.e., the poster) in the annotation decision. As such, ED models are expected to learn this feature from the data to achieve good performance.

Finally, Figure 1 illustrates the ambiguity of the event triggers in SuicideED by presenting the label distribution of the top 5 most frequent event trigger words. As can be seen, there is no particular dominant label for even the most frequent words. Hence, a ED model must effectively capture surrounding context of triggers to perform classification.

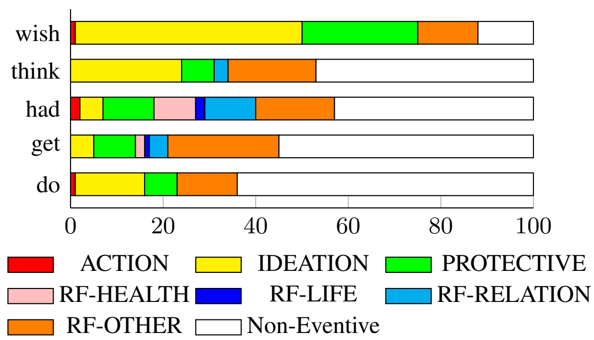

Figure 1: Label distribution of common trigger words.

4 Experiments

To assess the complexity of the ED task in SuicideED, we evaluate the performance of the following state-of-the-art ED models: **CNN**: a convolutional neural network for ED (Nguyen and Grishman, 2015); **DMBERT**: a dynamic multi-pooling model based on BERT (Wang et al., 2019); **BERTED**: a BERT-based model augmented with multi-layer perception (Yang et al., 2019); **BERTGCN**: a graph convolutional network (GCN) based on dependency trees (Nguyen and Grishman, 2018); **GatedGCN**: a GCN model using BERT and trigger-aware gating mechanism (Lai et al., 2020); and **EEGCN**: a GCN model that exploits syntactic structure and typed dependency information (Cui et al., 2020). All of the models leverage pre-trained BERT model to obtain representation vectors. The hyperparameters of the models are fine-tuned over the development data; a reproducibility checklist is presented in Appendix A. Finally, we further fine-tune the pre-trained BERT model over unlabeled Reddit posts (i.e., about 40K posts) using masked language modeling (Devlin et al., 2019). We report the model performance when the fine-tuned BERT replaces the original pre-trained BERT to explore the effectiveness of domain customization of BERT for the informal texts in Reddit.

Model	**BERT-base-cased**			**Finetuned BERT**		
	P	R	F	P	R	F
CNN	47.5	44.9	46.2	48.6	46.7	47.6
DMBERT	51.7	62.1	56.4	52.1	64.1	57.5
BERTED	47.8	66.3	55.5	48.8	65.3	55.8
BERTGCN	56.0	61.9	58.8	55.5	63.5	59.2
GatedGCN	54.6	64.1	59.0	54.2	65.1	59.2
EEGCN	54.6	65.5	59.5	53.7	66.7	59.5

Table 3: Performance of the models on the SuicideED test set using BERT and finetuned BERT embeddings.

Table 3 presents the performance of the models the SuicideED test set. The first observation is that fine-tuning BERT over Reddit posts can further improve the performance of the ED models although this is less pronounced for recent advanced ED models, i.e., GatedGCN and EEGCN. Second, the performance of the graph-based models (e.g., GatedGCN and EEGCN) is significantly better than those for non-graph-based models (i.e., CNN, DMBERT, and BERTED). As such, despite the informal nature of texts that can hinder the performance for dependency parsing, dependency trees are still helpful for the representation learning of ED models in SuicideED. Finally and most importantly, we find that the performance of existing ED models on SuicideED is substantially worse than the typical performance of such models on prior ED datasets (e.g., 77.6% on ACE-05 with GatedGCN and EEGCN) (Lai et al., 2020). It thus suggests the unique challenges of SuicideED for ED models and highlight the need for further research to improve ED for suicide-related events. Finally, we provide a statement for ethical issues in the fifth page.

5 Related Work

Suicide detection and prevention using NLP methods has caught the attention of many researchers. Due to the privacy restrictions associated with clinical databases, researchers have used publicly-available data from social media with manual annotations for recognizable signals of mental health issues (Coppersmith et al., 2015; Shing et al., 2018). The majority of methods, however, focus on detecting suicidal attempts or accessing suicide propensity of users based on social media posts (Coppersmith et al., 2015; Bhat and Goldman-Mellor, 2017; Shing et al., 2018; Zirikly et al., 2019). As such, these prior work has only relied the setting of overall text classification that fails to explore fine-grained analysis/classification at word level to reveal suicide-related events as we do.

Prior research effort for ED has introduced various methods for this problem, including feature engineering (Ahn, 2006; Ji and Grishman, 2008; Li et al., 2013) and deep learning (Chen et al., 2015; Wang et al., 2019; Cui et al., 2020) models. However, such prior work mainly utilizes the ED datasets with general event types and formal texts, i.e., ACE-05 (Walker et al., 2006), that might not be helpful for specific domains. Recently, there have been some effort on creating new datasets for ED in more specific domains, including biomedical texts (Kim et al., 2009), literary texts (Sims et al., 2019), and cybersecurity texts (Satyapanich et al., 2020). However, none of existing ED datasets explore suicide-related events in social media texts.

6 Conclusion

We present SuicideED, the first dataset focused on the event detection task for suicide-related events. SuicideED is manually annotated for 7 event types and provides enough training examples to develop large-scale deep learning models. We perform extensive evaluations of state-of-the-art ED models to demonstrate the challenges of the dataset and call for further effort to improve performance. In the future, we plan to extend SuicideED to annotate event arguments and other event properties better support event analysis and understanding for suicide.

7 Statement of Ethics and Human Subject Research

Working with sensitive data such as mental health information from human subjects requires taking special care. This becomes particularly relevant in this case as our main objective is to provide a dataset for general public use. Benton et al. (2017) discuss, however, that research with human subjects information is exempted from the required full Institutional Review Board (IRB) review if the data is already available from public sources or if the identity of the subjects cannot be recovered.

By design, Reddit is a platform where users remain anonymous and make their posts available to the general public. Nonetheless, additional privacy measures were taken by removing any username mentions from the documents as they can sometimes include identifiable information. Furthermore, unlike previous works where the main objective is to assess suicidal risk at the user level (Coppersmith et al., 2015; Bhat and Goldman-Mellor, 2017; Shing et al., 2018; Zirikly et al., 2019), this work focuses on sentence-level ED. As such, our dataset does not include any user-level information that could be used to identify individual subjects. Hence, this work is considered exempt from review by our University's IRB as the documents used are already publicly available and the original posters are impossible to identify.

Minimizing impact on the annotators: All prospective annotators were informed beforehand about the nature of the related text material and were made aware of its potential impact on their mental health. All chosen annotators had background knowledge/training on the subject at hand and were either clinicians or psychology graduates. Any candidates who reported suffering from, or having a history of, mental health-related issues were not considered out of concern for their health.

References

David Ahn. 2006. The stages of event extraction. In *Proceedings of the Workshop on Annotating and Reasoning about Time and Events*, pages 1–8.

Adrian Benton, Glen Coppersmith, and Mark Dredze. 2017. Ethical research protocols for social media health research. In *Proceedings of the First ACL Workshop on Ethics in Natural Language Processing*.

Harish S Bhat and Sidra J Goldman-Mellor. 2017. Predicting adolescent suicide attempts with neural networks. *NIPS 2017 Workshop on Machine Learning for Health.*

David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022.

Yubo Chen, Liheng Xu, Kang Liu, Daojian Zeng, and Jun Zhao. 2015. Event extraction via dynamic multi-pooling convolutional neural networks. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 167–176.

Glen Coppersmith, Ryan Leary, Eric Whyne, and Tony Wood. 2015. Quantifying suicidal ideation via language usage on social media. In *Joint Statistics Meetings Proceedings, Statistical Computing Section, JSM*, volume 110.

Shiyao Cui, Bowen Yu, Tingwen Liu, Zhenyu Zhang, Xuebin Wang, and Jinqiao Shi. 2020. Edge-enhanced graph convolution networks for event detection with syntactic relation. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2329–2339.

Corine de Ruiter and Tonia Nicholls. 2011. Protective factors in forensic mental health. In *International Journal of Forensic Mental Health*.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186.

Peter M. Gutierrez. 2006. Integratively assessing risk and protective factors for adolescent suicide. In *Suicide and Life-Threatening Behavior*.

Heng Ji and Ralph Grishman. 2008. Refining event extraction through cross-document inference. In *Proceedings of ACL-08: HLT*, pages 254–262.

Shaoxiong Ji, Celina Ping Yu, Sai-fu Fung, Shirui Pan, and Guodong Long. 2018. Supervised learning for suicidal ideation detection in online user content. *Complexity*, 2018.

Jin-Dong Kim, Tomoko Ohta, Sampo Pyysalo, Yoshinobu Kano, and Jun'ichi Tsujii. 2009. Overview of BioNLP'09 shared task on event extraction. In *Proceedings of the BioNLP 2009 Workshop Companion Volume for Shared Task*.

Viet Dac Lai, Tuan Ngo Nguyen, and Thien Huu Nguyen. 2020. Event detection: Gate diversity and syntactic importance scores for graph convolution neural networks. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 5405–5411.

Qi Li, Heng Ji, and Liang Huang. 2013. Joint event extraction via structured prediction with global features. In *Proceedings of the 51th Annual Meeting of the Association for Computational Linguistics*.

David N. Milne, Glen Pink, Ben Hachey, and Rafael A. Calvo. 2016. CLPsych 2016 shared task: Triaging content in online peer-support forums. In *Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology*, pages 118–127.

Thien Huu Nguyen and Ralph Grishman. 2015. Event detection and domain adaptation with convolutional neural networks. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, pages 365–371.

Thien Huu Nguyen and Ralph Grishman. 2018. Graph convolutional networks with argument-aware pooling for event detection. In *Thirty-second AAAI conference on artificial intelligence*.

Elizabeth O'Connor, Bradley Gaynes, Brittany Burda, Clara Williams, and Evelyn Whitlock. 2013. Screening for suicide risk in primary care: A systematic evidence review for the u.s. preventive services task force. In *Agency for Healthcare Research and Quality, U.S. Department of Health and Human Services*.

Joanne Ross, Shane Darke, Erin Kelly, and Kate Hetherington. 2012. Suicide risk assessment practices: A national survey of generalist drug and alcohol residential rehabilitation services. In *Drug and Alcohol Review*.

Taneeya Satyapanich, Francis Ferraro, and Tim Finin. 2020. Casie: Extracting cybersecurity event information from text. In *Proceedings of the AAAI conference on artificial intelligence*.

Han-Chin Shing, Suraj Nair, Ayah Zirikly, Meir Friedenberg, Hal Daumé III, and Philip Resnik. 2018. Expert, crowdsourced, and machine assessment of suicide risk via online postings. In *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*, pages 25–36.

Matthew Sims, Jong Ho Park, and David Bamman. 2019. Literary event detection. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3623–3634, Florence, Italy. Association for Computational Linguistics.

Hieu Duc Man Trong, Duc Trong Le, Amir Pouran Ben Veyseh, Thuat Nguyen, and Thien Huu Nguyen. 2020. Introducing a new dataset for event detection in cybersecurity texts. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 5381–5390, Online. Association for Computational Linguistics.

M Johnson Vioules, Bilel Moulahi, Jérôme Azé, and Sandra Bringay. 2018. Detection of suicide-related posts in twitter data streams. *IBM Journal of Research and Development*, 62(1):7–1.

Christopher Walker, Stephanie Strassel, Julie Medero, and Kazuaki Maeda. 2006. Ace 2005 multilingual training corpus. In *Technical report, Linguistic Data Consortium*.

Xiaozhi Wang, Xu Han, Zhiyuan Liu, Maosong Sun, and Peng Li. 2019. Adversarial training for weakly supervised event detection. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 998–1008.

Xiaozhi Wang, Ziqi Wang, Xu Han, Wangyi Jiang, Rong Han, Zhiyuan Liu, Juanzi Li, Peng Li, Yankai Lin, and Jie Zhou. 2020. MAVEN: A Massive General Domain Event Detection Dataset. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1652–1671, Online. Association for Computational Linguistics.

Sen Yang, Dawei Feng, Linbo Qiao, Zhigang Kan, and Dongsheng Li. 2019. Exploring pre-trained language models for event extraction and generation. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5284–5294.

Ayah Zirikly, Philip Resnik, Özlem Uzuner, and Kristy Hollingshead. 2019. CLPsych 2019 shared task: Predicting the degree of suicide risk in Reddit posts. In *Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology*, pages 24–33.

A Reproducibility checklist

- **Dataset**: The statistics of the created dataset SuicideED (including training/development/test portions) and the annotation process are presented in Section 2. The dataset is included in the submission. We will publicly release the dataset upon the acceptance of the paper. A URL to the publishing site will be included in the paper.

- **Source code with the specification of all dependencies, including external libraries**: We will publicly release the code to run the models upon the acceptance of the paper.

- **Description of computing infrastructure used**: All the experiments were run on a machine with 2 Intel Xeon E5-2620 v4 CPUs, 128GB of RAM, and 4 NVIDIA RTX 2080 Ti GPUs with 11GB RAM. We only train the models with one GPU. The amount of GPU memory for each run ranges from 5 to 7 GB, depending on the models being used.

- **Average runtime for each approach**: We train the models for 100 epochs; each takes approximately 2 minutes. The best epoch is chosen based on the performance on the development set.

- **Number of parameters in the model**: Every model uses the pre-trained BERT model with non-trainable 110M parameters. The CNN, DMBERT, BERTED, BERTGCN, GatedGCN, EEGCN models have additional 20M, 250K, 80K, 10M, 10M, and 8M trainable parameters, respectively.

- **Explanation of evaluation metrics used, with links to code**: Follow prior work in ED (Nguyen and Grishman, 2015; Chen et al., 2015), we use the precision, recall, and F1 scores for performance metrics.

- **Hyperparameter bounds and configurations for best-performing models**: We use the bert-base-cased version of BERT in all the considered models (Devlin et al., 2019). To obtain the representation vector for a trigger word candidate in a sentence, the hidden vectors of 12 layers of BERT are concatenated. We fine-tune the hyper-parameters for the models in this work over the development data of SuicideED. As such, to train the models, we use the Adam optimizer with the learning rate of $2e-5$ (searched in the range of $\{2e-5, 3e-5, 4e-5, 5e-5\}$) and batch size of 128 (searched in the range of $\{32, 64, 128, 256\}$). For the CNN model, we use 4 kernel sizes of $2, 3, 4, 5$, each with 150 filters (searched in the range $\{100, 150, 200, 250, 300\}$. The BERTGCN, GatedGCN, EEGCN models employ two GCN layers (searched in the range $\{2, 3, 4, 5\}$), each with 256 hidden units (searched in the range $\{128, 256, 512\}$). The edge embedding size of the EEGCN model is set to 50. Finally, we use two layers for all the feed-forward neural networks in the models with 256 hidden units in the layers (searched in the range $\{128, 256, 512\}$).

B Topic Modeling

To better understand the topics related to suicide in the SuicideED dataset, we run a topic modeling analysis using Latent Dirichlet Allocation (LDA) (Blei et al., 2003) over the documents in the dataset. We extract ten topics from the analysis and present their words in Table 4. English stop-words, the least ($p < 0.01$), and most frequent words ($p > 0.2$) were removed in the analysis. Interestingly, it can be observed that posts can be summarized into 3 main categories: school (2, 8), work (5, 9, 10), and family (3, 4, 6, 7), which somehow reflects the sources of mental issues.

#	Words
1	hate, thoughts, point, stop, care, say, worse, living
2	wish, hate, worse, try, year, shit, school, thoughts
3	world, parents, suicidal, right, bad, point, person
4	tell, love, tired, days, person, doing, death, mom
5	pain, real, tried, need, maybe, work, hurt, tired, talk
6	care, told, tired, said, parents, bad, need, leave, right
7	care, matter, getting, days, actually, feels, parents
8	got, school, said, talk, doing, self, love, mental, work
9	job, happy, got, year, love, hate, try, told, money
10	love, shit, job, work, suicidal, night, pain, right, year

Table 4: Topic models with LDA

C Annotation Guideline

Table 5 and 6 present a detailed description of event types and examples for each event type in our SuicideED dataset.

Type	Description	Examples
ACTION	This category includes any event of an individual engaging in actions that bring them closer to dying by suicide. These include any previous suicide attempts, preparatory acts towards a future attempt, or self-inflicted violence. When annotating this type of event, it is important that an actual action takes place and that it goes beyond verbalization or intent. As such, sentences containing these events mainly talk about the past or ongoing situations.	A previous suicide attempt is a self-inflicted, potentially injurious behavior with an intent to die as a result. *I tried to **kill** myself last night.*
		A preparatory act consists of any acts of preparation toward making a suicide attempt. Must be beyond verbalization or thought such as assembling a method (e.g. buying a gun, collecting pills) or preparing for death (e.g. writing a suicide note and a will). *Just **looked** online for the quickest way* *I **left** a note for my parents*
		Self-inflicted violence includes self-directed, harmful behaviors that do not have a clear intent to die as a result. *I've started **cutting** myself again*
IDEATION	These events focus on expressing thoughts and feelings but no actual action is present. These, however, are not related to actions such as preparatory acts and, instead, refer to verbalizations of inner feelings/desires.	It includes passive thoughts about wanting to be dead: *I **wish** I was dead* And, active thoughts about killing oneself. *I am going to **kill** myself soon*
PROTECTIVE	These events are related to capacities, qualities, environmental and personal resources that increase resilience; drive an individual toward growth, stability, health, and/or an increase in coping with different life events. For this category, please annotate any sentence that showcases a positive impact on an individual. These can be verbalizations of self-worth and willingness to get better, access to medical resources, positive personal relationships, positive cultural beliefs, etc.	Access to effective behavioral health care and medication: *My therapist says that I should **talk** more.* *The medication seems to be **helping**.*
		Connectedness to individuals, family, community, and social institutions: *At least my **friends** are there for me*
		Life skills (including problem-solving skills and coping skills, ability to adapt to change): *I've always been **good** at helping people*
		Self-esteem and a sense of purpose or meaning in life: *My life is much **better** than many people.*
		Expressing a willingness to improve: *I really **want** to get better.* *I **wanna** be funny and outgoing.*
		Cultural, religious, or personal beliefs that discourage suicide: *I know God **disapproves** of what I'm thinking.*

Table 5: Event types with their descriptions and examples in the SuicideED dataset. Event trigger words are shown in bold. Continued in Table 6

.

Type	Description	Examples
RF - LIFE	This risk factor event is easy to identify as a loss of life of a both human an non-human entities. The loss of life might be explicitly or implicitly expressed.	Loss of a relative, explicitly expressed: *After my brother **killed** himself...*
		Loss of a relative, implicitly expressed: *My grandma has been **gone** for years now.*
		Loss of a pet friend: *My dog just **died**, he was my only real friend.*
RF - RELATIONSHIP	These include events such as social isolation, family breakdowns, divorce, etc. Include in this category all events that show a loss of connection with other people. These can be both verbalization of feelings of isolation or actual incidents of loss of an interpersonal relationship such as a break-up or argument with another individual.	Social isolation: *I don't **have** anyone to talk to.*
		Family breakdown: *My dad just **kicked** me out of the house.*
		Divorce: *After my **divorce**, I started drinking*
RF - HEALTH	These include events such as social isolation, family breakdowns, divorce, etc. Include in this category all events that show a loss of connection with other people. These can be both verbalization of feelings of isolation or actual incidents of loss of an interpersonal relationship such as a break-up or argument with another individual.	Mental disease/disorder such as depression, PTSD: *Can't **deal** with my depression right now.* *I've been **diagnosed** with BPD.*
		Chronic or long-term disease, pain, and disability: *I'm just **giving** into my eating disorder.* *Recently, my diabetes has been **acting** up.*
		Misuse and abuse of alcohol or other drugs: *I've been **drinking** a lot lately*
RF - OTHER	These events include all other risk factors that do not fall into the LIFE, RELATIONSHIP, or HEALTH categories. As such, these can be events of very diverse natures such as financial issues, chronic abuse, discrimination, or general quality of life problems.	Financial hardship: *Can't **afford** to pay rent anymore...*
		Prison: *I can't go back to **jail** now.*
		Job loss: ***Lost** my job today.*
		Discrimination *They **tease** me in school cause I'm gay*

Table 6: Event types with their descriptions and examples in the SuicideED dataset. Event trigger words are shown in bold.