Learning Peer Support Interactions via Bi-LSTM Graph Neural Networks for Suicide Risk Prediction

Anonymous ACL submission

Abstract

Suicide prevention through early detection using social media data has been widely studied. However, the critical role of peer support—interactions among individuals with similar mental disorders—has not been deeply investigated or exploited. In this study, we explore peer interactions in online communities for individuals with bipolar disorder and leverage this information to predict suicide risk levels. We propose a model that uses contextualized posts and comments along with their sentiment features. By embedding these features into a peer support network, our model captures peer interactions and predicts suicide risk levels using bidirectional LSTM Graph Neural Networks (Bi-LSTM GNNs). Experimental results demonstrate the effectiveness of our approach, outperforming baseline methods. Our findings highlight the importance of peer comments in predicting suicide risk.

1 Introduction

The global rise in social media usage has created new avenues for interaction, significantly impacting suicide detection and prevention efforts (Intahchomphoo, 2018; Leiva and Freire, 2017). Social media interactions can influence users' mental states, as they often disclose sensitive information and receive either positive or negative feedback from peers. This interaction can affect their posting behavior and potentially escalate suicidal tendencies or aid in recovery (De Choudhury et al., 2016). Especially, people with mental disorders are more impacted by peer support, which involves interactions among individuals with similar mental disorders (Shalaby and Agyapong, 2020).

Previous studies have demonstrated the potential of using social media activity to predict users' mental health. For instance, (Cao et al., 2019) predicted suicide risk by creating suicide-oriented embeddings with layered attention, and (Lee et al., 2023) analyzed users' timeline sequences on Reddit to predict suicidality. Additionally, (Sawhney et al., 2022) developed a Hyperbolic Conversation Network to detect suicide ideation through online conversations. However, the effect of peer support on social media has not been extensively explored.

Therefore, in this paper, we present a novel model to predict suicide risk by comprehending peer support in online mental disorder communities. Specifically, we utilize the dataset from a study by (Lee et al., 2023), which includes posts from users diagnosed with bipolar disorder across three mental health-related subreddits, along with their comments. It is well known that bipolar patients have a higher risk of suicide compared to those with other psychiatric or medical conditions (Rihmer and Kiss, 2002; Ilgen et al., 2010). Depressive episodes in mood disorders are a major contributor to increased mortality due to suicide (Jamison, 2019).

To understand peer support relationships, we create a peer support graph with posts and comments. The posts and comment nodes are connected based on user interactions, incorporating time sequence information. We then embed contextualized posts and comments, along with their sentiment, as node features. Next, we propose a Bi-LSTM GNN-based model that can learn distinct peer support characteristics from posts with a high risk of suicide. By adopting Bi-LSTM, our model captures the dynamics of post context and user interactions, while the GNN learns the structural relationships among peers.

Experimental results show that our model accurately predicts suicide risk, demonstrating the importance of peer support interactions for suicide prevention. This method holds potential for future applications in predicting and preventing suicide based on users' interactions and support systems.

Table 1: The number of posts and comments written by 560 bipolar patients in each suicide risk level

Labels	Posts	Comments
Indication	4,198	12,430
Ideation	679	1,436
Behaviour	156	510
Attempt	67	124
Total	5,100	14,500

2 Related Work

Social media platforms have become convenient spaces for expressing thoughts that individuals may find difficult to share in person. While acknowledging the negative impacts of social media on mental health, several studies have highlighted its potential benefits. For instance, (Lee et al., 2022) utilized graph construction based on Reddit posts to predict users' suicide risks. (Abdulsalam and Alhothali, 2022) proposed a combined LSTM-Attention-CNN model to detect suicidal intentions from social media submissions. (Renjith et al., 2022) explored various data types across multiple social platforms, developing machine learning models to identify users at risk. (Cao et al., 2020) constructed a suicide-oriented knowledge graph using deep neural networks for detecting suicidal ideation in social media posts. (Hussain et al., 2020) predicted users' mental health statuses based on their posts and comments, they did not focus specifically on suicide detection.

However, peer support, which is recognized as an important factor in mental health communities, has not yet been thoroughly explored. Addressing this gap, we propose a context-aware peer support learning model using a bi-LSTM Graph Neural Network approach.

3 Dataset

To leverage peer support information among individuals with mental disorders in our model, we utilize the Bipolar Reddit dataset (Lee et al., 2023). The dataset contains 7,692 posts published between January1st, 2008, and September 30th, 2021, from the three representative bipolar-related subreddits, including `r/bipolar`, `r/BipolarReddit`, and `r/BipolarSOs`. These posts are authored by 818 users who are clinically diagnosed with bipolar disorder and are annotated by two clinical psychiatrists into four levels of suicide risk: *Indication*, *Ideation*, *Behavior*, and *Attempt*.

Since the original dataset only includes posts, we proceeded to collect comments on all these posts to explore peer support within mental disorder communities. Note that we excluded users whose accounts no longer exist and kept 560 remaining users for additional data collection. Using the open-source *Reddit* API[1], we fetched 211,920 comments corresponding to 5,100 Reddit posts. In our research context, we focus on utilizing 14,500 comments specifically written by 560 users, who are considered as peers within the community. The number of posts and comments written by those users are provided in Table 1.

4 Methodology

4.1 Problem Statement

In this paper, we aim to predict the suicide risk level of users at the post level by leveraging temporal post features along with peer support information. Suppose we have a set of posts $P = \{p_i\}_{i=1}^{|P|}$ and a set of comments $C = \{c_j^i\}_{j=1}^{|C|}$ which correspond to each post p_i. Then, the proposed model is defined as a multi-class classification problem, where each post p is classified into one of four suicide risk levels $y_i \in \{Indicator, Ideation, Behaviour, Attempt\}$.

4.2 Peer Support Graph

Our proposed model leverages the benefits of graph structures to capture relationships among individuals with mental disorders and their social media posts. In the following subsections, we describe the construction of a peer support graph and the embedding of node features into the graph.

4.2.1 Graph Construction

In order to consider temporal sequence of posts and peer support relationships, we construct a heterogeneous graph that comprises post and comment nodes. Let $G = (V, E)$ be a graph where V is the set of vertices composed of posts P and comments C. The set of edges E is composed of the three types of edges to represent as follows; (i) **Post-Comment Relationship**: For each post p_i, there exists an edge (p_i, c_j^i) for each comment c_j^i associated with p_i; (ii) **Temporal Sequence of Posts**: Posts authored by the same user are connected sequentially in the order of their timestamps. Specifically, a post node p_i is connected to its predecessor p_{i-1} and its successor p_{i+1} if such posts exist,

[1] https://www.reddit.com/dev/api/

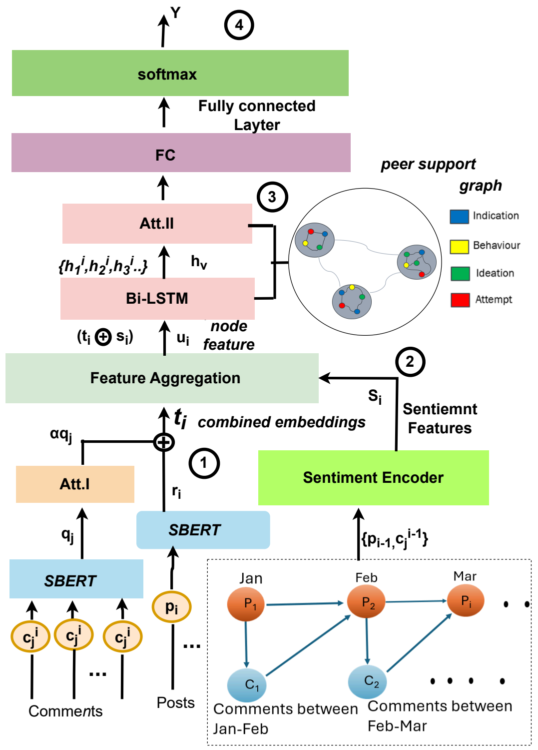

Figure 1: The overall architecture of the proposed model: ① post-comment embedding relationship , ② Sentiment Analysis Layer, ③ Bi-LSTMGraphsage Attention Layer, ④ Suicide Risk Prediction.

forming edges (p_{i-1}, p_i) and (p_i, p_{i+1}); (iii) **Peer Support Relationship**: To represent peer support relationships, a comment node c_j is connected to a post node p_h, where p_h is one of the posts authored by the user who wrote c_j. The post p_h is chosen such that the timestamp of p_h is closest to the timestamp of c_j, creating edges (c_j, p_h).

4.2.2 Node Features

We next embed node features into the peer support graph by incorporating (i) contextualized posts and comments, and (ii) sentiment features.

Contextualized Post/Comment Encoder: Our proposed method encodes posts and comments to capture their meaning and context. The contextualized encoder adopts the pre-trained SBERT (Reimers and Gurevych, 2019), a language model that demonstrates robust performance across various NLP tasks. After tokenizing each post and comment into the set of m and n words, respectively, we encode them as follows.

$$r_i = \text{SBERT}(p_i) \in \mathbb{R}^{m \times d_t} \quad (1)$$

$$q_j = \text{SBERT}(c_j) \in \mathbb{R}^{n \times d_t} \quad (2)$$

where d_t is the dimension of the textual feature. We then add self-attentive comment embeddings to the corresponding post node feature. This approach helps to place more weight on the more important comments related to the post. The final contextualized post embedding t_i is calculated as follows.

$$\alpha_j = \frac{\exp(\mathbf{a}^T q_j)}{\sum_k \exp(\mathbf{a}^T q_k)} \quad (3)$$

$$t_i = r_i \oplus \sum_j \alpha_j q_j \quad (4)$$

Sentiment Encoder: Understanding sentiment changes over time is crucial, as it might indicate a high risk of suicide, especially in individuals with bipolar disorder. For example, a bipolar patient who expresses negative emotions in a post and does not receive positive and supportive comments might experience an increased risk of suicide in the future. Therefore, our proposed model incorporates a sentiment encoder that generates sentiment embeddings from the previous post and its corresponding comments. The sentiment encoder incorporates the fine-tuned DistilBERT (Sanh et al., 2019) with the Stanford Sentiment Treebank (SST) (Socher et al., 2013). The sentiment embedding, s_i, of post p_i is the sum of the sentiment of the previous post and the average sentiment of comments from the previous post. This is calculated as follows.

$$s_i = f(p_{i-1}) \oplus \frac{\sum_{c_j \in p_{i-1}} f(c_j)}{|\{c_j\}|} \quad (5)$$

where $f(\cdot)$ is the fine-tuned DistilBERT.

The final post representation is then an aggregation of contextualized and sentiment features.

$$u_i = t_i \oplus s_i \quad (6)$$

4.3 Bi-LSTM Graph Neural Network

Our proposed model leverages Bi-LSTM Graph Neural Networks (GNN) to effectively integrate both structural and sequential information from the peer support graph. Bidirectional LSTM is employed to capture the temporal dynamics of posts and comments by processing sequences bidirectionally, which is essential for discerning the evolution of sentiments and contextual meanings over time in social media interactions. In conjunction, GraphSAGE (Hamilton et al., 2017) is utilized within the GNN framework to handle large-scale graphs efficiently by sampling and aggregating features from local neighborhoods. By combining Bi-LSTM with GraphSAGE, our model enhances the generation of node embeddings, enabling them to encompass richer representations that encapsulate both temporal dynamics and relational contexts.

We first initialize node embedding, $\mathbf{h}_v^0$, with the node features. We then apply Bi-LSTM to the sampled neighbors, $\mathcal{N}_v^k$, as follows.

$$\overrightarrow{\mathbf{h}_v} = \overrightarrow{\text{LSTM}}\left(\{\mathbf{h}_u^k \mid u \in \mathcal{N}_v^k\}\right) \tag{7}$$

$$\overleftarrow{\mathbf{h}_v} = \overleftarrow{\text{LSTM}}\left(\{\mathbf{h}_u^k \mid u \in \mathcal{N}_v^k\}\right) \tag{8}$$

$$\mathbf{h}_v = [\overrightarrow{\mathbf{h}_v}; \overleftarrow{\mathbf{h}_v}], \tag{9}$$

To prioritize relevant post information, we apply attention (Vaswani et al., 2017) as follows.

$$\alpha_{ij} = \frac{\exp(\mathbf{a}^T[\mathbf{h}_i^{(k)} \| \mathbf{h}_j^{(k)}])}{\sum_{k \in \mathcal{N}(i)} \exp(\mathbf{a}^T[\mathbf{h}_i^{(k)} \| \mathbf{h}_k^{(k)}])} \tag{10}$$

The node representations at $k+1$ iterations and final representation are calculated as follows.

$$\mathbf{h}_i^{(k+1)} = \sigma\left(\sum_{j \in \mathcal{N}(i)} \alpha_{ij} \mathbf{W} \mathbf{h}_j^{(k)}\right) \tag{11}$$

$$\mathbf{h}_v^{\text{final}} = \mathbf{z}_v^K \tag{12}$$

Finally, a prediction vector is generated for a given post p_v.

$$\hat{y} = \mathcal{F}(ReLU(\mathcal{F}(p_v))) \tag{13}$$

where $\mathcal{F}$ is a fully-connected layer and $ReLU$ is an activation function.

5 Experiment

To evaluate the performance of our proposed model, we conduct experiments with the following baseline methods; Suicide Detection Model (SDM) (Cao et al., 2019), CNN (Shing et al., 2018), Context-GNN (Lee et al., 2022), MTL (Lee et al., 2023), HGAT (Wang et al., 2019), and STATENet (Sawhney et al., 2020).

As shown in Table 2, our proposed model outperforms all baseline methods. This suggests that our approach facilitates a comprehensive understanding of user behaviors and interactions in mental health contexts, supporting more accurate predictions, by leveraging the bidirectional-LSTM GNN on the peer support graph.

Table 2: Suicide risk prediction results of the proposed model and the baseline models

Model	Metrics			
	Accuracy	Precision	Recall	F1
SDM	0.72	0.71	0.75	0.73
CNN	0.75	0.76	0.78	0.76
STATEnet	0.69	0.65	0.72	0.68
Context-GNN	0.77	0.76	0.80	0.78
HGAT	0.73	0.74	0.81	0.77
MTL	0.82	0.79	**0.86**	0.81
Ours	**0.83**	**0.80**	0.85	**0.83**

Table 3: Performance of proposed model variations by excluding specific components

Component	Precision	Recall	F1
Full model	0.80	0.85	0.83
(−) Attention	0.76	0.80	0.79
(−) Bi-LSTM	0.79	0.74	0.75
(−) GNN	0.72	0.77	0.73
(−) Comment	0.64	0.55	0.61

5.1 Ablation Study

We perform an ablation study to assess the impact of each component of our proposed model, as detailed in Table 3. Our findings indicate that every component plays a crucial role in predicting suicide risk. Notably, excluding attention mechanisms, Bi-LSTM, and GNN results in a moderate decrease in performance. However, the most substantial performance drop occurs when comment information is omitted, underscoring the importance of understanding peer interactions in predicting suicide risk.

6 Conclusion

In this study, we proposed a comprehensive model for predicting suicide risk by leveraging peer support interactions in mental health communities. Our approach integrates a bidirectional LSTM-based GNN designed to capture the temporal dynamics of contextualized posts and comments, while also modeling the relationships among peers. Our findings highlight the pivotal role of understanding peer comments in effective suicide risk prediction. The model demonstrates considerable clinical adaptability for anticipating and preventing suicidality. We intend to publish our code to facilitate reproducibility and further advancements in this critical area of research.

Limitations

This study has several limitations that need to be addressed in future research:

First, our work is limited to bipolar disorder communities, as the dataset we used exclusively contains data from bipolar patients. To address this limitation, we plan to apply our model to other datasets that include various mental disorder communities. Additionally, we aim to expand our data collection efforts to include a broader range of mental health conditions.

Second, our proposed model currently lacks a focus on interpretability. Understanding the reasons behind the model's decisions is crucial for clinicians, as it aids in diagnosis and treatment planning. Therefore, we plan to develop an additional component to enhance the interpretability of our model.

Finally, our study is restricted to mental disorder-related subreddits. Many individuals with mental disorders may also be active in other subreddits unrelated to mental health. To gain a comprehensive understanding of their social media activities, we intend to expand our data collection to include a wider range of subreddits and online communities.

Ethical Statement

The dataset utilized in this study (Lee et al., 2023) adheres to IRB and ACL ethical considerations. To ensure user privacy, personally identifiable information such as usernames and IDs in the additional comments collected has been substituted with unique identifiers, thus anonymizing user identities. Our procedures strictly comply with our institution's ethical policies regarding data handling and user confidentiality.

References

Asma Abdulsalam and Areej Alhothali. 2022. Suicidal ideation detection on social media: A review of machine learning methods. *arXiv preprint arXiv:2201.10515*.

Lei Cao, Huijun Zhang, and Ling Feng. 2020. Building and using personal knowledge graph to improve suicidal ideation detection on social media. *IEEE Transactions on Multimedia*, 24:87–102.

Lei Cao, Huijun Zhang, Ling Feng, Zihan Wei, Xin Wang, Ningyun Li, and Xiaohao He. 2019. Latent suicide risk detection on microblog via suicide-oriented word embeddings and layered attention. *arXiv preprint arXiv:1910.12038*.

Munmun De Choudhury, Emre Kiciman, Mark Dredze, Glen Coppersmith, and Mrinal Kumar. 2016. Discovering shifts to suicidal ideation from mental health content in social media. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, pages 2098–2110.

Will Hamilton, Zhitao Ying, and Jure Leskovec. 2017. Inductive representation learning on large graphs. *Advances in neural information processing systems*, 30.

Shahid Hussain, Anam Nasir, Khurram Aslam, Subhan Tariq, Mian Farhan Ullah, et al. 2020. Predicting mental illness using social media posts and comments. *International Journal of Advanced Computer Science and Applications*, 11(12).

Mark A Ilgen, Amy SB Bohnert, Rosalinda V Ignacio, John F McCarthy, Marcia M Valenstein, H Myra Kim, and Frederic C Blow. 2010. Psychiatric diagnoses and risk of suicide in veterans. *Archives of general psychiatry*, 67(11):1152–1158.

Channarong Intahchomphoo. 2018. Social media and youth suicide: A systematic review.

Kay Redfield Jamison. 2019. Suicide and bipolar disorder. *Bipolar Disorder*, pages 115–119.

Daeun Lee, Migyeong Kang, Minji Kim, and Jinyoung Han. 2022. Detecting suicidality with a contextual graph neural network. In *Proceedings of the eighth workshop on computational linguistics and clinical psychology*, pages 116–125.

Daeun Lee, Sejung Son, Hyolim Jeon, Seungbae Kim, and Jinyoung Han. 2023. Towards suicide prevention from bipolar disorder with temporal symptom-aware multitask learning. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 4357–4369.

Victor Leiva and Ana Freire. 2017. Towards suicide prevention: early detection of depression on social media. In *Internet Science: 4th International Conference, INSCI 2017, Thessaloniki, Greece, November 22-24, 2017, Proceedings 4*, pages 428–436. Springer.

Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*.

Shini Renjith, Annie Abraham, Surya B Jyothi, Lekshmi Chandran, and Jincy Thomson. 2022. An ensemble deep learning technique for detecting suicidal ideation from posts in social media platforms. *Journal of King Saud University-Computer and Information Sciences*, 34(10):9564–9575.

Zoltán Rihmer and Kitty Kiss. 2002. Bipolar disorders and suicidal behaviour. *Bipolar Disorders*, 4:21–25.

Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2019. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*.

Ramit Sawhney, Shivam Agarwal, Atula Tejaswi Neerkaje, Nikolaos Aletras, Preslav Nakov, and Lucie Flek. 2022. Towards suicide ideation detection through online conversational context. In *Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval*, pages 1716–1727.

Ramit Sawhney, Harshit Joshi, Saumya Gandhi, and Rajiv Shah. 2020. A time-aware transformer based model for suicide ideation detection on social media. In *Proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP)*, pages 7685–7697.

Reham A Hameed Shalaby and Vincent IO Agyapong. 2020. Peer support in mental health: literature review. *JMIR mental health*, 7(6):e15572.

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.

Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D Manning, Andrew Y Ng, and Christopher Potts. 2013. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 conference on empirical methods in natural language processing*, pages 1631–1642.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Xiao Wang, Houye Ji, Chuan Shi, Bai Wang, Yanfang Ye, Peng Cui, and Philip S Yu. 2019. Heterogeneous graph attention network. In *The world wide web conference*, pages 2022–2032.

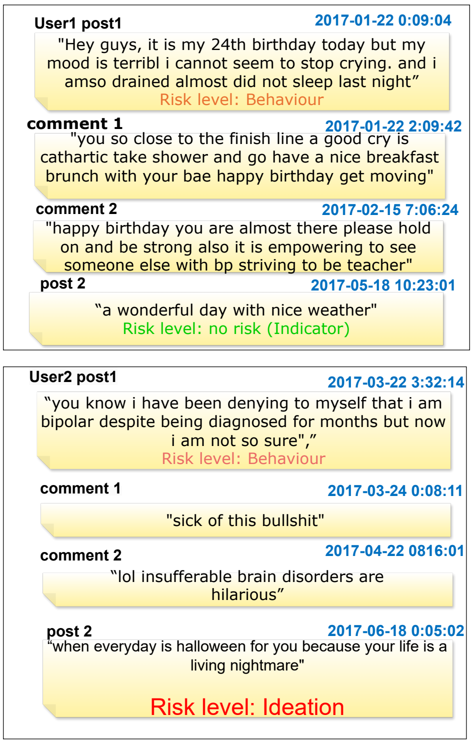

Figure 2: Example Reddit posts along with supportive comments from peers.

A Appendix

A.1 Experiment Setting

We split the dataset into train and test sets with 80-20 ratio. All the experiments were performed in the AMD Threadripper Pro 3955WX system integrated with a NVIDIA RTX 4090.

A.2 Case Study

Figure 2 presents example posts along with their corresponding comments, highlighting peer support interactions on Reddit. These examples demonstrate how the risk of suicide appeared to decrease after the individuals received supportive comments from their peers.