Proceedings of the Eighteenth International AAAI Conference on Web and Social Media (ICWSM 2024)

# Leveraging Psychiatric Scale for Suicide Risk Detection on Social Media

**Bichen Wang, Pengfei Deng, Song Chen, Yanyan Zhao***, **Bing Qin**

Harbin Institute of Technology, Heilongjiang, China
{bichenwang,pfdeng,songchen,yyzhao,qinb}@ir.hit.edu.cn

## Abstract

The objective of suicide risk detection on social media is to identify individuals who may attempt suicide and determine their suicide risk level based on their online behavior. Although data-driven learning models have been used to predict suicide risk levels, these models often lack theoretical support and explanation from psychiatric research. To address this issue, we propose the incorporation of professional psychiatric scales into research to provide theoretical support and explanations for our model. Our proposed Scale-based Neural Network (SNN) architecture aims to extract content associated with scales from the posting history of social media users to predict their suicide risk level. Additionally, our approach provides scale-based explanations for the model's predictions. Experimental results demonstrate that our proposed method outperforms several strong baseline methods and highlights the potential of combining psychiatric scales and computational techniques to improve suicide risk detection.

## Introduction

Suicide is a major global public health concern, ranking among the top twenty leading causes of death worldwide. According to the World Health Organization, approximately 800,000 people lose their lives to suicide annually (Organization et al. 2019). The recent COVID-19 pandemic has further exacerbated this issue, resulting in a surge in suicide rates attributed to economic downturns and social unrest (Sher 2020; Chan, Sahimi, and binti Mokhzani 2022; Torjesen 2020; John et al. 2020). Consequently, governments and healthcare organizations worldwide have made preventing suicide a top priority. The study of automatic suicide risk detection is an essential tool to tackle this issue.

Social media provides an outlet for individuals with suicidal ideation to express their thoughts and intentions, which they may keep hidden from their loved ones (Park, McDonald, and Cha 2021; Coppersmith et al. 2018; De Choudhury et al. 2013). Computational methods offer an avenue to automatically identify groups at risk of suicide on social media. Specifically, suicide risk detection involves classifying users into five risk levels based on their posts within a specified time frame, thereby framing it as a classification problem.

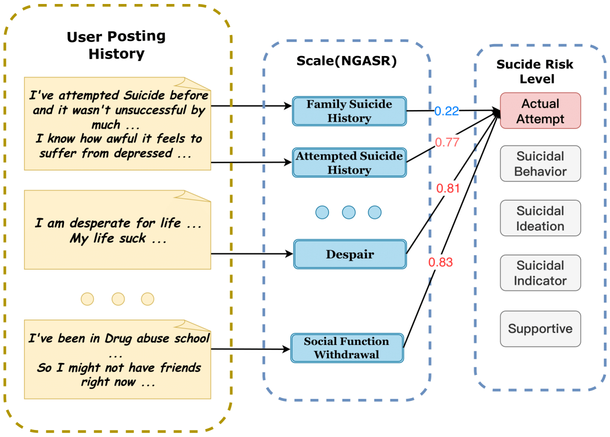
User Posting History
I've attempted Suicide before and it wasn't unsuccessful by much ...
I know how awful it feels to suffer from depressed ...
I am desperate for life ...
My life suck ...
I've been in Drug abuse school ...
So I might not have friends right now ...
Scale(NGASR)
Family Suicide History
Attempted Suicide History
Despair
Social Function Withdrawal
0.22
0.77
0.81
0.83
Sucide Risk Level
Actual Attempt
Suicidal Behavior
Suicidal Ideation
Suicidal Indicator
Supportive

Figure 1: An illustration of our suicide risk detection. The model employs posts associated with scale items and measures their scores to arrive at a decision.

Several methods have been proposed for detecting suicide risk on social media, using both traditional machine learning and deep learning techniques (Sawhney et al. 2018; Gaur et al. 2019; Sawhney et al. 2021a; Zirikly et al. 2019). Previous research has primarily focused on improving the performance of suicide risk detection models. Nevertheless, these models often lack explainability and professional theoretical foundations, which renders them often similar to a black box, and consequently raises challenges for users, such as psychiatrists, to trust and accept the generated alerts.

To enhance our model's reliability and explainability, we propose integrating traditional psychiatric research findings with emerging computational techniques. Specifically, our computational model is guided by the use of professional psychiatric scales. In our study, we utilize the Nurse Global Assessment of Suicide Risk (NGASR) scale, which is a scale widely used by psychiatric professionals because of its relative objectivity and ease of use (Walsh M. 1988). The NGASR comprises 15 items that psychiatric professionals assess and assign scores to determine an individual's suicide risk level. Figure 1 illustrates how our model simulates this

[*]* Corresponding author

process by associating each item on the scale with a user's post. Ultimately, the model predicts the score by integrating all relevant posts for each item. By considering all items in the scale, the model determines their suicide risk level. For instance, a user's post such as "I might not have any friends" might relate to the item "Social Function Withdrawal". Finally, the model aggregates the results from each item to predict the user's suicide risk level.

To this end, we propose using a Scale-based Neural Network (SNN) with three stages: scale items representation, suicide evidence matching, and risk level prediction, to predict an individual's suicide risk level by modeling the various items in the scale. The stage of scale items representation requires using template sentences to represent items in NGASR, and we use a prototyping learning method to enhance the quality of their representations. After the scale items representation, we evaluate each item based on the user's posts and score them accordingly. Eventually, we predict the individual's suicide risk level based on the evaluation outcomes obtained during the suicide evidence matching stage. The integration of a professional psychiatric scale improves the performance and explainability of our model. We conduct extensive experiments to demonstrate the superiority of our model over strong baseline models.

Our contributions are as follows:

- We propose integrating findings from conventional psychiatric research with modern computational techniques applied to social media and suggest the inclusion of a professional psychiatric scale to detect suicide risk in social media.
- We propose a novel model, the SNN, that integrates traditional psychiatric research scales with emerging computational technologies and investigate its performance.
- We conduct a series of experiments that demonstrated the advantages of SNN. Our experiments highlight the meaningful benefits of integrating computational techniques with traditional psychiatric tools.

## Related Work

In this section, we review the related work in two interconnected yet distinct subdomains: psychological problem detection on social media, and suicide risk detection. While suicide risk detection is a specific and critical aspect of psychological problem detection, it demands a unique set of approaches and techniques due to the complex nature and severity of the issue.

### Psychological Problem Detection On Social Media

We discuss the use of social media data analysis in the study of prevalent psychological issues, such as depression, anxiety, schizophrenia, and more. Contemporary people like to share their lives on social media. At the same time, this information reveals their inner activity and self-character, which can provide very little information to the model. Recently, researchers have carried out a large number of studies analyzing users' psychology on social media using computing techniques. It has been proposed that social media data can be used to detect a variety of mental disorders, such as depression, PTSD, and anxiety (Choudhury et al. 2013; Coppersmith et al. 2015). Many researchers have attempted to detect depression among social media users. Depression detection has seen improvements through the use of both Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), which have independently yielded more accurate results than traditional methods  (Husseini Orabi et al. 2018). The CLPsych-2015 task challenge proposed the detection of post-traumatic stress disorder (PTSD) and depression by participating teams  (Coppersmith et al. 2015). Given the long posting history of users and the amount of useless information contained, a summary approach and reinforcement learning methods have been used to extract the most relevant information about depression (Zogan et al. 2021; Gui et al. 2019). To provide a more fine-grained classification of depression, four levels (non-depression, mild, moderate, and severe) have been proposed instead of just two (non-depression and depression) (Naseem et al. 2022). Psychological studies have shown that users with depression exhibit specific behaviors on social media, so both user behaviors and posts are combined as features (Zogan et al. 2022). The language model has also been used to detect anxiety disorders in Twitter users (Owen, Camacho-Collados, and Anke 2020).

### Suicide Risk Detection

Researchers have employed data from diverse sources and various statistical learning techniques to detect suicide risk. Some studies have explored the prediction of patients' suicide risk through professional electronic health records (EHRs) (Rawat et al. 2022). Suicide notes have been analyzed to assess the impact of loneliness and despair on suicide (Ghosh, Ekbal, and Bhattacharyya 2022). The social media suicide risk dataset has been constructed by collecting Reddit posts and having them labeled by psychologists (Shing et al. 2018). Adhering to psychological standards, researchers have classified suicide risk into five levels and compared the performance of different models (Gaur et al. 2019). Considering the vast amount of unlabeled data online, researchers have employed unsupervised learning to categorize information related to potential suicide risk on social media (Parraga-Alava et al. 2019). During the shared task Clpsych-2019, the best results are achieved by using pre-trained models and psychological features (Zirikly et al. 2019). Ordinal loss functions have been proposed as an alternative to cross-entropy loss functions for training models (Sawhney et al. 2021a). With the advancement of graph neural networks, social graph networks have been suggested to investigate the spread of suicidal thoughts on social media (Sawhney et al. 2021b, 2022). Several researchers currently investigate techniques for identifying and incorporating psychological characteristics linked to suicide, such as emotional stability, mental anxiety, negative life events, and stress, among others (Sawhney et al. 2021a; Guzman-Nateras et al. 2022; De Choudhury and Kiciman 2017; Lee et al. 2022).

Our work introduces a comprehensive suicide risk scale, a direct and widely-accepted professional tool. This tool dif-

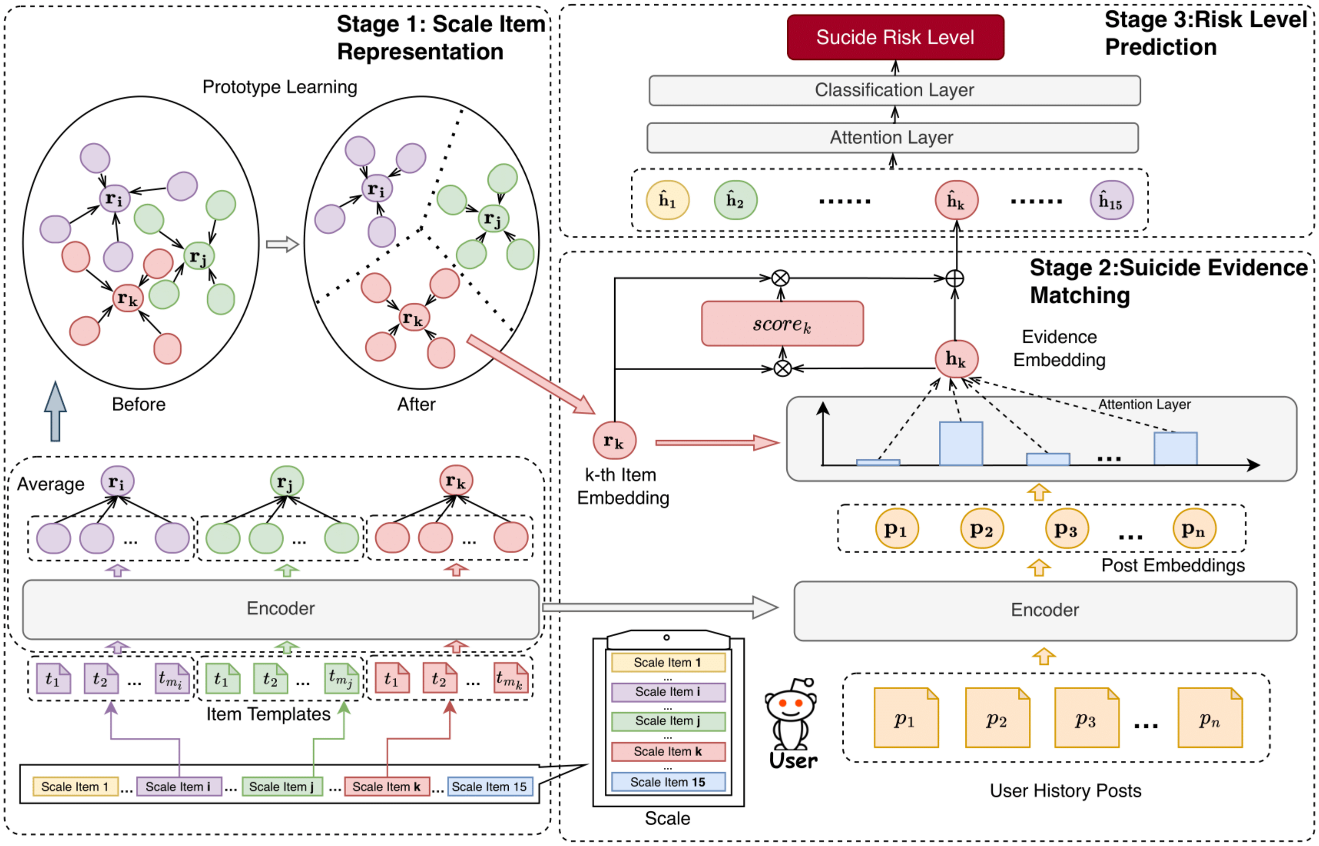
Stage 1: Scale Item Representation
Prototype Learning
$\mathbf{r_i}$
$\mathbf{r_j}$
$\mathbf{r_k}$
Before
After
Average
$\mathbf{r_i}$
$\mathbf{r_j}$
$\mathbf{r_k}$
...
Encoder
$t_1$ $t_2$ ... $t_{m_i}$
$t_1$ $t_2$ ... $t_{m_j}$
$t_1$ $t_2$ ... $t_{m_k}$
Item Templates
Scale Item 1 ... Scale Item i ... Scale Item j ... Scale Item k ... Scale Item 15
Stage 3:Risk Level Prediction
Sucide Risk Level
Classification Layer
Attention Layer
$\hat{\mathbf{h}}_1$ $\hat{\mathbf{h}}_2$ ...... $\hat{\mathbf{h}}_k$ ...... $\hat{\mathbf{h}}_{15}$
Stage 2:Suicide Evidence Matching
$score_k$
$\mathbf{h_k}$
Evidence Embedding
$\mathbf{r_k}$
k-th Item Embedding
Attention Layer
...
$\mathbf{p_1}$ $\mathbf{p_2}$ $\mathbf{p_3}$ ... $\mathbf{p_n}$
Post Embeddings
Encoder
Scale Item 1
Scale Item i
Scale Item j
Scale Item k
Scale Item 15
Scale
User
$p_1$ $p_2$ $p_3$ ... $p_n$
User History Posts

Figure 2: SNN's overview. The SNN consists of three stages, namely scale item representation, suicide evidence matching, and risk level prediction.

ferentiates our work from previous studies. Compared to other works, the scale we introduced is more direct and widely used and accepted by experts.

## Three-stage Scale-based Neural Network

For a user $U_i$ who has made posts $P_i = p_{i1}, \cdots, p_{iL}$, where $p_{ij}$ represents the $j$-th post generated by $U_i$, the goal of suicide risk detection is to predict the suicide risk level, denoted by $\hat{y} \in [0, 4]$. In this study, we categorize users into one of five classes based on their suicide risk level: Supportive (SU), Suicidal Indicator (IN), Suicidal Ideation (SI), Suicidal Behavior (BR), and Actual Attempt (AT).

The SNN architecture, as depicted in Figure 2, consists of three stages: scale item representation, suicide evidence matching, and risk level prediction. This model takes the user's posting history as input. Firstly, the scale item representation module encodes items in psychiatric scales into item embeddings. Next, the suicide evidence matching module utilizes these item embeddings to extract relevant posts related to each item in the psychiatric scales and scores them accordingly. Finally, the risk level prediction module leverages the output of the previous stage to predict the suicide risk level.

### Scale Item Representation

As shown in Table 1, based on the content reflected in the scale, there are a total of 15 scale items in NGASR. In this section, we aim to represent them using item embeddings, which will later be used to search for evidence of suicide risk in a user's posting history.

**Item Templates** Each item in the scale is presented as a simple and clear phrase, which is convenient for experts to use, but unsuitable for modeling. As shown in Table 1, each scale item is actually just an indicative noun in the scale, such as "Psychiatric History", which may offer very little information to the model. To address this issue, we have developed a collection of 300 concise and straightforward template sentences, based on psychiatric expertise, to represent 15 different items (denoted as $t_i$). These templates are not derived from the Reddit dataset or any other data sources. Instead, they have been curated based on psychiatric expertise. Each template sentence is a statement that captures the essence of the item in the scale. For instance, the template sentence "I have attempted suicide" corresponds to the "History of Suicide Attempts" item, while "I am experiencing severe depression" represents the "Psychiatric History" item. Examples of different items and their corresponding templates are provided in Table 1.

| Scale Item | No. of templates | Template Example |
|---|---|---|
| Despair | 20 | I feel hopeless about life. |
| Hallucinations | 20 | I have horrible voices in my head. |
| Recent Negative Life Events | 20 | I am unemployed , unfortunately. |
| Loss of Interest | 20 | I am not intere sted in anything. |
| Social Function Withdrawal | 20 | I avoid people around me. |
| Family History of Suicide | 20 | My father commi tted suicide. |
| Losing Close Relationship | 20 | I lost my best friend. |
| Psychiatric History | 20 | I suffer from d epression. |
| History of Suicide Attempts | 20 | I have tried suic ide. |
| Suicide Plan | 20 | I have a suicid e gun ready. |
| Alcohol and Drugs | 20 | I drink and smo ke marijuana a lot. |
| Low Socioe-conomic Status | 20 | I am a poor man. |
| Advanced Disease | 20 | I have cancer a nd I'm dying. |
| Talking about Suicide | 20 | Suicide can be a beautiful thing. |
| Widow or Widower | 20 | I am a widow. |
| **All** | 300 | - |

Table 1: 15 scale items in NGASR and their templates

Subsequently, we focus to represent each item using the templates.

**Scale Item Representation By Prototype Learning** To effectively represent each item in the scale using a limited number of templates, our approach aims to represent them in a meaningful way. Although averaging all template sentence embeddings from the BERT model is a straightforward representation approach based on templates, it has limitations. Directly averaging these embeddings could lead to information loss and inadequate differentiation among items, posing challenges in achieving representative item embeddings for the scale.

To overcome these challenges, we finetune the BERT model before the formal training of our model and adopt a prototype learning approach. Prototype learning is a metric-based technique designed specifically for few-shot learning, which allows the network to create representative prototypes for each class. As depicted in Figure 2, the prototype learning process results in a more concentrated distribution of template sentence embeddings around the item embeddings, improving the distinguishability of the item embeddings. By adopting this approach, BERT can generate more distinct and meaningful embeddings for each item. The mathematical representation of this procedure is as follows:

$$\mathbf{r_k} = \frac{1}{|T_k|}\sum_{t_i \in T_k} f_\phi(t_i) \tag{1}$$

where $T_k$ refers to the set of templates for the $k$-th scale item, $t_i$ denotes the $i$-th template sentence in $T_k$, and $f_\phi$ represents the BERT model. The resulting $\mathbf{r_k}$ represents the item embedding of the $k$-th item. We will not delve into the specific steps of prototype learning as described in other papers (Snell, Swersky, and Zemel 2017); instead, we present the loss function of prototype learning. The specific loss function is as follows:

$$l = -\sum_{k=1}^{15}\sum_{t_i \in T_k} \log\left(\frac{\exp(\rho(f_\phi(t_i), \mathbf{r_k}))}{\sum_{\hat{k}} \exp(\rho(f_\phi(t_i), \mathbf{r_{\hat{k}}}))}\right) \tag{2}$$

where the distance function used is the cosine distance function, denoted as $\rho$. Using prototype learning, this approach of representing scale items is anticipated to produce more precise and distinguishable item embeddings.

## Suicide Evidence Matching

At this stage, the SNN identifies posts related to the scale items and provides a score for each item. The SNN extracts posts from the user's social media history that are directly related to the scale items and attempts to assign a score to each item, thereby enhancing confidence in the reliability of our model.

**Evidence for Scale Items** The aim of this section is to collect evidence closely related to the scale to support the detection of suicidal risk in users. The evidence here is posts directly linked to the scale items, which can verify the presence of several items in the scale. Figure 2 displays the post embeddings produced by a BERT model after prototype learning. BERT generates post embeddings by transforming each user's post $p$ into a post embedding $\mathbf{p}$ using the following function:

$$\mathbf{p} = f_\phi(p) \tag{3}$$

To associate each post with the 15 scale items, we use an attention mechanism that calculates the weighted sum of the retrieved post embeddings with the scale item embeddings to generate an evidence embedding for each scale item. This mechanism also determines the contribution of each post to each scale item. Specifically, we calculate the attention score $\alpha_{ij}$ for each post embedding $\mathbf{p_j}$ and scale item embedding $\mathbf{r_i}$ using the equation:

$$e_{ij} = \frac{\mathbf{r_i}^\top \cdot \mathbf{p_j}}{\sqrt{d}} \tag{4}$$

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_k \exp(e_{ik})} \tag{5}$$

$$\mathbf{h_i} = \sum_j \alpha_{ij} * \mathbf{p_j} \tag{6}$$

where, $\mathbf{h_i}$ represents the evidence embedding that is obtained by the weighted summation of all post embeddings related to the scale item $\mathbf{r_i}$, and $d$ denotes the dimension of $\mathbf{r_i}$ and $\mathbf{p_j}$. The attention score $\alpha_{ij}$ specifies the contribution of the $j$-th post embedding to the evidence embedding for the $i$-th scale item.

The attention score enables us to identify posts directly related to the scale. Additionally, we employ a function to score each item on the scales, which is based on evidence embeddings $\mathbf{h}$ extracted from users' social media posting histories. We analyze all posts and combine the evidence embeddings from the posts with the item embeddings to score each item.

**Scores for Scale Items** The aim of this section is to integrate the scale item with users' posting history and assign a score to each item. To facilitate human understanding, the scores should be independent of each other and presented within a specific range. To calculate the scores, we choose the cosine similarity function since it returns values within the range of $[-1, 1]$.

As illustrated in Figure 2, we apply the cosine similarity function, which ensure that the scores of one scale item do not influence others, and multiple items can be severe at the same time. The computation is detailed in the following equation:

$$score_i = \frac{\mathbf{r_i}^\top \cdot \mathbf{h_i}}{\|\mathbf{r_i}\| * \|\mathbf{h_i}\|} \tag{7}$$

In addition to facilitating explainability, the $score$ can aid us in achieving a more refined representation. To derive the final deep representation of the user and scale, we employ the following equation:

$$\hat{\mathbf{h}}_\mathbf{i} = Relu(\mathbf{W}(\mathbf{h_i} \cdot score_i + \mathbf{r_i}) + \mathbf{b}) \tag{8}$$

where, $\mathbf{W} \in R^{h \times d}$ and $\mathbf{b} \in R^h$ are learnable parameters, and $h$ is the dimension of $\hat{\mathbf{h}}_\mathbf{i}$. We add $\mathbf{r_i}$ and $score_i \cdot \mathbf{h_i}$ to obtain the further representation $\hat{\mathbf{h}}_\mathbf{i}$. The item embedding $\mathbf{r}$ is directly involved in the representation $\hat{\mathbf{h}}_\mathbf{i}$.

Our model is capable of identifying posts that are relevant to specific scale items, determining a score for each item, and it enhances transparency, comprehensiveness, and reliability by employing a scale-based method, thus improving upon previous models. Finally, the model predicts the level of suicide risk by aggregating the information collected during the assessment.

## Risk Level Prediction

Having obtained the evidence and scores of each scale item, our model is able to predict a user's suicide risk level. At this stage, the model employs an attention mechanism to generate the final representation using the following equations:

$$\beta_i = \frac{\exp(\mathbf{v}^\top \cdot \hat{\mathbf{h}}_\mathbf{i})}{\sum_{j=1}^{15} \exp(\mathbf{v}^\top \cdot \hat{\mathbf{h}}_\mathbf{j})} \tag{9}$$

$$\hat{\mathbf{h}}_\mathbf{all} = \sum_{i=1}^{15} \beta_i \cdot \hat{\mathbf{h}}_\mathbf{i} \tag{10}$$

where $\mathbf{v} \in R^h$ represents a learnable parameter employed to weigh all representations.

The attention mechanism weighs the summation of $\hat{\mathbf{h}}_\mathbf{1}, \cdots, \hat{\mathbf{h}}_\mathbf{15}$, which is subsequently fed into a feedforward neural network (FFN) for classification:

$$\hat{\mathbf{y}} = FFN(\hat{\mathbf{h}}_\mathbf{all}) \tag{11}$$

where $\hat{\mathbf{y}} \in R^5$, and $\hat{y}_i$ denotes the predicted probability of the label being $i$. The goal is to minimize the cross-entropy error for each user, in comparison with the ground-truth label $\mathbf{y}$:

$$l = -\sum_{i=0}^{4} y_i \cdot \log\left(\frac{\exp(\hat{y}_i)}{\sum_{j=1}^{5} \exp(\hat{y}_j)}\right) \tag{12}$$

This model predicts suicide risk levels, enabling psychiatric professionals to validate the $\hat{\mathbf{y}}$ values predicted by the model.

# Experimental Setups

## Datasets

We use a Reddit dataset (Gaur et al. 2019), which consists of the posting history of 500 users collected from Reddit. Four practitioner psychiatrists perform annotation following the guidelines of the C-SSRS (Posner et al. 2008).

The entire dataset is classified into five user categories. According to psychological research, the suicide process can be classified into four stages: the emergence of suicidal triggers, suicidal ideation, possible suicidal actions, and the decision to commit suicide. Accordingly, the dataset has four categories: Suicidal Indicator (IN), Suicidal Ideation (SI), Suicidal Behavior (BR), and Actual Attempt (AT). The users who are considered to be at no risk but likely to offer support to those with suicidal tendencies on Reddit are labeled as Supportive (SU) in the dataset. The dataset distribution is presented in Table 2.

| Suicide Risk Class | No. of User | Avg. No. of posts | Avg. No. of words |
|---|---|---|---|
| Supportive (SU) | 108 | 20.00 | 67.72 |
| Suicidal Indicator(IN) | 99 | 17.18 | 72.21 |
| Suicidal Ideation(SI) | 171 | 24.55 | 73.81 |
| Suicidal Behavior(BR) | 77 | 19.70 | 63.56 |
| Actual Attempt(AT) | 45 | 17.81 | 85.90 |

Table 2: Statistical characteristics of the dataset. Each class is named according to psychology.

The data is stratified into a ratio of 4:1, resulting in a training set of 400 users and a testing set of 100 users. To ensure the robustness of our experiments, we perform 5-fold holdout cross-validation on the train set, where each fold comprises 80 users. The test set is initially held out and used for final evaluation. To account for the variance in results, we conduct multiple runs of each experiment, initializing the random seed differently each time. Our experiments are repeated 20 times, and we report the average of the results obtained.

## Evaluation Metrics

For the evaluation metrics, the traditional evaluation metrics are modified when the datasets and tasks are proposed (Gaur et al. 2019). They alter the formulation of False Negatives (FN) and False Positives (FP). Meanwhile, Graded Precision and Graded Recall are presented. FN is modified as the ratio of the number of times the predicted level of $k^p$ is less than the actual risk level $k^a$ over the size of test data $N_T$. FP is the ratio of the number of times the predicted risk $k^p$ is greater than the actual risk $k^a$.

$$FN = \frac{\sum_{i=1}^{N_T} I(k_i^a > k_i^p)}{N_T} \tag{13}$$

$$FP = \frac{\sum_{i=1}^{N_T} I(k_i^a < k_i^p)}{N_T} \tag{14}$$

The Graded Precision and Graded Recall use the newly defined FP and FN.

## Implementation Details

We use the following hyperparameters: a hidden state size of 256 features and a dropout rate of 0.7. Firstly, we conduct prototype learning on the BERT language model to ensure reliable generation of item embeddings. The learning rate for this task is set to 3e-4. Next, during overall tuning training, we fine-tune the language model with a learning rate of 1e-6, while other components are fine-tuned with a learning rate of 1e-4. We implement all methods using PyTorch 1.10 and optimize them with AdamW (Loshchilov and Hutter 2019) using a batch size of 8 and weight-decay of 1e-3. To handle user histories of different lengths, we store post-embedding sequences as packed padding sequences. We trained the model on Nvidia A100 GPUs for 50 epochs and applied early stopping with a patience of 5 epochs.

## Comparison Methods

We compare our approach with several existing methods, including those based on traditional neural networks and pretrained models. Many of these methods have been previously employed by researchers.

- **Contextual CNN** (Gaur et al. 2019): This method utilizes Convolutional Neural Networks (CNN) to address the task of suicide detection, as suggested by prior studies.
- **BiLSTM+Attention** (Sawhney et al. 2021a): This model employs Bidirectional Long Short-Term Memory (BiLSTM) to capture the post sequence and incorporates an attention mechanism to focus on specific posts for suicide risk detection.
- **BERT+Max Pooling**: This model encodes the posts using BERT and applies max pooling directly.
- **BERT+Average Pooling**: Similar to the above model, this approach employs BERT encoding and performs average pooling directly.
- **ContextBERT** (Matero et al. 2019): The ContextBERT model achieves the best performance in the task of detecting suicide risk at CLPsych 2019. It incorporates a large number of psychological user features to detect suicide risk. It utilizes BERT to encode Reddit posts and models the user's posting sequence and user features using Gated Recurrent Units (GRU).
- **SISMO** (Sawhney et al. 2021a): The SISMO model utilizes the differences in suicide risk level labels to design an ordinal loss function and encode the posts using Longformer (Beltagy, Peters, and Cohan 2020).
- **SASI** (Sawhney, Neerkaje, and Gaur 2022): The SASI model incorporates a risk avoidance mechanism, enabling it to refrain from making predictions when uncertainty arises regarding suicide risk.

To ensure consistency in our comparison, all models, except for Contextual CNN and SISMO, are encoded using BERT. The Glove word vector (Pennington, Socher, and Manning 2014) is used in the CNN model, while Longformer is employed in SISMO.

# Results

## Comparison Result

As shown in Table 3, results demonstrate that our method outperforms all the competitive baselines in the suicide risk detection task. In addition to overall performance improvement, as illustrated in Figure 3, the confusion matrix highlights the effectiveness of the SNN model in identifying high-risk users.

| Model | Graded Recall | Graded Precision | F-Score |
|---|---|---|---|
| Contextual CNN | 0.52 | **0.69** | 0.59 |
| BiLSTM+Attn | 0.57 | 0.63 | 0.60 |
| Bert+Avgpooling | 0.59 | 0.56 | 0.57 |
| Bert+Maxpooling | 0.54 | 0.57 | 0.55 |
| ContextBERT | 0.59 | 0.61 | 0.60 |
| SISMO | 0.61 | 0.66 | 0.64 |
| SASI | 0.62 | 0.67 | 0.66 |
| **SNN** | **0.73*** | 0.68* | **0.70*** |

Table 3: Main results of the experiments. The best results have been bolded. * indicates result is statistically significant compared to SASI.

**Overall Performance** As depicted in Table 3, our model outperforms all baseline models in terms of recall and F-Score. Our model's explicit incorporation of scales enables more accurate identification of users with different suicide risk levels, reducing prediction errors. Although our model ranks second in terms of precision, the difference between our model's precision and the top-performing model is very small. These results demonstrate the effectiveness of our explicit modeling of scales in detecting users at risk for suicide.

**High-Risk Detection** In real-world applications, individuals actively involved in suicidal behavior (BR) or those attempting suicide (AT) are categorized as *high-risk* users, while individuals displaying suicidal ideation (SI) are classified as *mid-risk* users. All remaining individuals (IN, SU) are deemed *low-risk* users. Precisely identifying high-risk users is vital in real-world applications, as detecting more high-risk individuals can potentially save lives. Thus, we examine our models' capacity to distinguish between different levels of suicide risk.

Figure 3 illustrates a comparative analysis of the confusion matrix obtained from our proposed model and three other well-performing models. The experimental results clearly indicate that the remaining models exhibits limited proficiency in identifying high-risk users, particularly when it comes to detecting individuals attempting suicide (AT). In contrast, our model demonstrates significant efficacy in accurately detecting high-risk users. Our model enhances the identification of both AT and BR users. SNN is able to differentiate subtle differences in various suicide risk levels, while remaining models hardly recognize high-risk user especially AT users. The introduction of psychiatric scales has greatly aided the model in better distinguishing different levels of suicide risk.

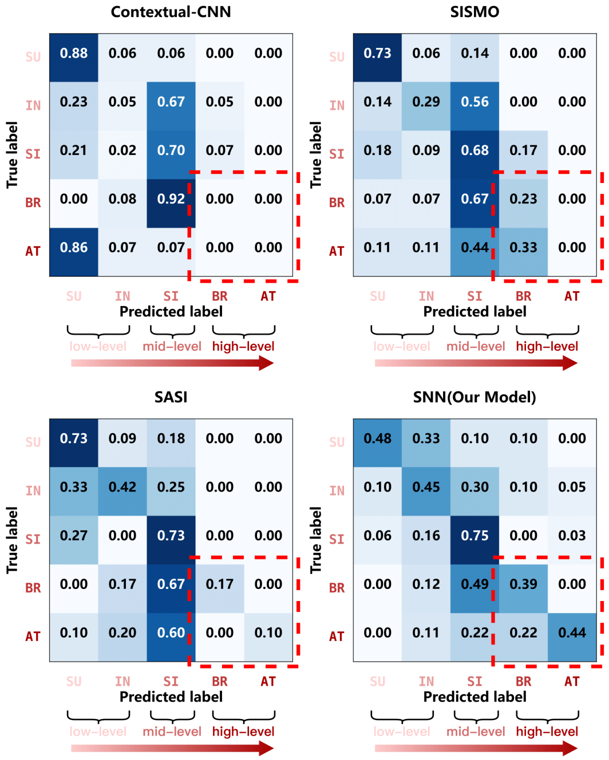

Figure 3: The figure illustrates the normalized confusion matrix, where the rows represent the actual labels of suicide risk levels, and the columns represent the predicted suicide risk levels generated by the model. The diagonal elements display the percentage of correct predictions. The high-risk user predictions are highlighted with a red box.

## Ablation Study

Some ablation experiments are conducted to analyze the contribution of each part of the model. The experiments are conducted for several purposes, including the validation of the necessity of item embeddings and suicide evidence matching. The contribution of each part of the model is analyzed to gain insights into how it affects the overall performance of the model.

- −scale item embedding: We use average pooling template embeddings to replace item embeddings
- −scale item templates: Instead of using item embeddings obtained from multiple templates, we utilize a single one, five, or ten templates.
- −suicide evidence matching: We do not assign a score to each item on the scale. Instead, we directly feed each evidence embedding into the final risk level prediction.

| Model | Graded Recall | Graded Precision | F-Score |
|---|---|---|---|
| **SNN** | **0.73** | **0.68** | **0.70** |
| −scale item embedding | 0.66 | 0.63 | 0.65 |
| −scale item templates(half) | 0.70 | 0.64 | 0.67 |
| −scale item templates(quarter) | 0.64 | 0.59 | 0.62 |
| −scale item templates(single) | 0.63 | 0.57 | 0.60 |
| −suicide evidence matching | 0.68 | 0.66 | 0.67 |

Table 4: Results of our ablation experiments

Table 4 shows the results of our ablation experiment. The findings indicate that item embeddings are superior to average pooling template embeddings. Moreover, the results of using a single template are similar to the method without templates at all. As the number of templates increases, the model performance continues to improve. We believe that if more templates are provided, our model will obtain a stronger representation of item semantics and thus enable further performance improvements. This indicates that our scale contributes to the model's ability to detect suicide risk. Incorporating suicide evidence matching enhances the interaction between item embeddings and evidence embeddings, improving the model's ability while also providing a more detailed explanation of the results.

## Explainability Study

We believe our model will be able to provide its explanations based on attention mechanisms and scale scores. Due to the lack of consensus on the definition and understanding of explainability, evaluating explainability is challenging without standards. We attempt to study the explainability of our model through expert evaluation and case studies.

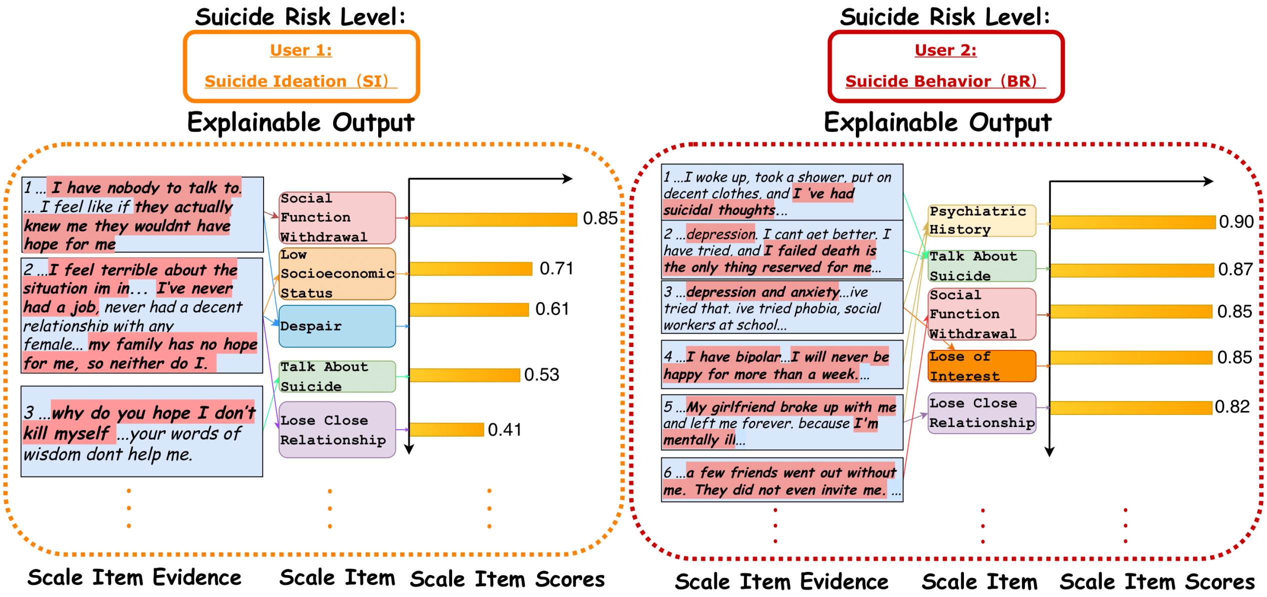
Suicide Risk Level:
User 1:
Suicide Ideation (SI)
Explainable Output
1 ... I have nobody to talk to. ... I feel like if they actually knew me they wouldnt have hope for me
2 ...I feel terrible about the situation im in... I've never had a job, never had a decent relationship with any female... my family has no hope for me, so neither do I.
3 ...why do you hope I don't kill myself ...your words of wisdom dont help me.
Social Function Withdrawal
Low Socioeconomic Status
Despair
Talk About Suicide
Lose Close Relationship
0.85
0.71
0.61
0.53
0.41
Scale Item Evidence
Scale Item
Scale Item Scores
Suicide Risk Level:
User 2:
Suicide Behavior (BR)
Explainable Output
1 ...I woke up, took a shower, put on decent clothes, and I 've had suicidal thoughts...
2 ...depression. I cant get better. I have tried, and I failed death is the only thing reserved for me...
3 ...depression and anxiety...ive tried that. ive tried phobia, social workers at school...
4 ...I have bipolar...I will never be happy for more than a week....
5 ...My girlfriend broke up with me and left me forever. because I'm mentally ill...
6 ...a few friends went out without me. They did not even invite me. ...
Psychiatric History
Talk About Suicide
Social Function Withdrawal
Lose of Interest
Lose Close Relationship
0.90
0.87
0.85
0.85
0.82
Scale Item Evidence
Scale Item
Scale Item Scores

Figure 4: Two cases of SNN predictions. SNN can be explained as a primary assessment of a user's suicide risk on social media. The sentences that could reflect suicide risk have been marked in red. We present the top five highest-scoring items and linked each post to its corresponding item. The histogram also displays the scores for each item.

**Matching Accuracy Evaluation** Two psychiatry experts were involved in evaluating the model's ability to match posts with scale items. The evaluation primarily focused on the top five ranked items generated by the model, alongside their corresponding posts that had the highest attention scores. The experts were provided with these posts and the corresponding item scores for analysis. The accuracy of the matching was validated by assessing the agreement between the experts and the model's accurate matching of the posts to the scale items. The experts reviewed 631 posts obtained from the posting history of 100 users and analyzed their corresponding scale items. If the experts determined the matches to be correct, the results were validated.

The averaged matching accuracy of our model was 79%, indicating that in 79% of cases, the model accurately associated the scale items with the posts ($Kappa = 0.87$). In the majority of cases, our model accurately established the correlation between the psychiatric scale and the user's posting history, offering plausible explanations. However, in some cases, the model might have made mistakes. Based on expert evaluation, we identified that a limited understanding of implicit semantics was the primary factor responsible for these limitations. To illustrate this, let's examine the following statement: ”Friend, please stop contacting these people. I know you don't want to be in touch with anyone anymore.” In this case, the model failed to recognize that the user was giving advice to someone and incorrectly associated it with ”Social Function Withdrawal”. It could be seen that modeling for scales is meaningful. In the future, we believe it will be necessary to add more templates for scale items.

## Case Study for Explainability

In order to demonstrate the ability of SNN to elucidate suicide risk detection outcomes, we present suicide risk results generated by our model using two cases. We show the top five rated scale items from the model output, as well as the posts with their highest attention scores. These are the factors that the model considers important. These cases serve to illustrate how we interpret the model's predictions.

As depicted in Figure 4, the model's suicide risk prediction is contingent upon the evidence of scale items and their scores. For instance, posts containing the phrase “I have bipolar disorder” from user-2 are considered as evidence for the scale item ”Psychiatric History.” Subsequently, the model assesses the score of “Psychiatric History” based on the entirety of the evidence for the scale item. It becomes evident that “Psychiatric History” constitutes a significant scale item, in alignment with our model's prediction. Ultimately, the model forecasts that user-2 belongs to the BR user group, which corresponds to the actual label.

By comparing user-1 and user-2, two cases representing different suicide risk levels, it becomes apparent that the model accurately discerns the distinctions between them. User-2 demonstrates scale items that are more prominent and contribute to a higher risk of suicide in comparison to user-1. Users at high risk, such as user-2, consistently have scale items that are more severe than users at low risk, like user-1, thus indicating that the explanation provided by our model is reliable.

The analysis of case studies demonstrates that our model possesses greater transparency and explainability compared to previous models. Our SNN model accurately associates scale items with relevant evidence posts and effectively determines the importance of these scale items. This enhanced transparency contributes to a more comprehensive understanding of the decision-making process of the model.

## Broader Impact and Ethical Considerations

This study aims to improve suicide risk detection on social media by incorporating a professional psychiatric scale into the proposed SNN model. While our research has the potential to provide valuable benefits to society, it also raises several ethical considerations and possible negative outcomes that need to be addressed.

### Ethical Considerations

The research on suicide risk may raise certain ethical concerns. The data used in this study are acquired from publicly shared datasets shared by other researchers. In order to protect individuals' privacy, all social media data underwent strict anonymization procedures by the data providers before being used. We comply with relevant ethical guidelines and legal regulations, ensuring that there is no risk of privacy violations during the research process. The classification is not intended as a diagnostic tool, but rather a risk estimate for individual users that can then be used to support monitoring and evidence-based prevention and support for users.

### Positive Outcomes

- Enhanced explainability and reliability: By incorporating a professional psychiatric scale, our model provides more transparent and comprehensible processes and results, thereby increasing trust and acceptance among professionals.
- Enhanced awareness of mental health: The development and implementation of our model have the potential to cultivate broader public awareness regarding mental health concerns and the significance of timely intervention for individuals at risk.
- Resource Allocation: The medical system can allocate resources more efficiently, ensuring that those at the highest risk receive attention immediately.

### Negative Outcomes and Mitigation Strategies

- False positives/negatives: Misclassification of individuals' suicide risk levels could lead to unnecessary interventions or missed opportunities for help. To address this, we continuously refine our model to improve its accuracy and emphasize that the model should be used as a supplementary tool, with final assessments made by mental health professionals.
- Clinical limitations: Despite the use of excellent clinical scales, further validation and verification are still required to ensure its accuracy and reliability in a clinical setting. Our research demonstrates the potential of combining computation with psychiatry or psychology, rather than advocating for the immediate deployment of the model for clinical use.

## Acknowledgments

We thank the anonymous reviewers for their insightful comments and suggestions. This work was supported by the National Key RD Program of China via grant 2021YFF0901602 and the National Natural Science Foundation of China (NSFC) via grants 62176078 and 62236004.

## References

Beltagy, I.; Peters, M. E.; and Cohan, A. 2020. Longformer: The long-document transformer. *arXiv preprint arXiv:2004.05150*.

Chan, L. F.; Sahimi, H. M. S.; and binti Mokhzani, A. R. 2022. A global call for action to prioritize healthcare worker suicide prevention during the CoViD-19 pandemic and beyond.

Choudhury, M. D.; Gamon, M.; Counts, S.; and Horvitz, E. 2013. Predicting Depression via Social Media. In Kiciman, E.; Ellison, N. B.; Hogan, B.; Resnick, P.; and Soboroff, I., eds., *Proceedings of the Seventh International Conference on Weblogs and Social Media, ICWSM 2013, Cambridge, Massachusetts, USA, July 8-11, 2013*. The AAAI Press.

Coppersmith, G.; Dredze, M.; Harman, C.; Hollingshead, K.; and Mitchell, M. 2015. CLPsych 2015 Shared Task: Depression and PTSD on Twitter. In *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, 31–39. Denver, Colorado: Association for Computational Linguistics.

Coppersmith, G.; Leary, R.; Crutchley, P.; and Fine, A. 2018. Natural language processing of social media as screening for suicide risk. *Biomedical informatics insights*, 10: 1178222618792860.

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

De Choudhury, M.; and Kiciman, E. 2017. The language of social support in social media and its effect on suicidal ideation risk. In *Proceedings of the international AAAI conference on web and social media*, volume 11, 32–41.

Gaur, M.; Alambo, A.; Sain, J. P.; Kursuncu, U.; Thirunarayan, K.; Kavuluru, R.; Sheth, A.; Welton, R.; and Pathak, J. 2019. Knowledge-aware assessment of severity of suicide risk for early intervention. In *The world wide web conference*, 514–525.

Ghosh, S.; Ekbal, A.; and Bhattacharyya, P. 2022. Am I No Good? Towards Detecting Perceived Burdensomeness and Thwarted Belongingness from Suicide Notes. In Raedt, L. D., ed., *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, 5073–5079. International Joint Conferences on Artificial Intelligence Organization. AI for Good.

Gui, T.; Zhu, L.; Zhang, Q.; Peng, M.; Zhou, X.; Ding, K.; and Chen, Z. 2019. Cooperative multimodal approach to depression detection in twitter. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, 110–117.

Guzman-Nateras, L.; Lai, V.; Veyseh, A. P. B.; Dernoncourt, F.; and Nguyen, T. 2022. Event detection for suicide understanding. In *Findings of the Association for Computational Linguistics: NAACL 2022*, 1952–1961.

Husseini Orabi, A.; Buddhitha, P.; Husseini Orabi, M.; and Inkpen, D. 2018. Deep Learning for Depression Detection of Twitter Users. In *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*, 88–97. New Orleans, LA: Association for Computational Linguistics.

John, A.; Pirkis, J.; Gunnell, D.; Appleby, L.; and Morrissey, J. 2020. Trends in suicide during the covid-19 pandemic.

Lee, D.; Kang, M.; Kim, M.; and Han, J. 2022. Detecting suicidality with a contextual graph neural network. In *Proceedings of the eighth workshop on computational linguistics and clinical psychology*, 116–125.

Loshchilov, I.; and Hutter, F. 2019. Decoupled Weight Decay Regularization. In *ICLR (Poster)*. OpenReview.net.

Matero, M.; Idnani, A.; Son, Y.; Giorgi, S.; Vu, H.; Zamani, M.; Limbachiya, P.; Guntuku, S. C.; and Schwartz, H. A. 2019. Suicide Risk Assessment with Multi-level Dual-Context Language and BERT. In *Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology*, 39–44. Minneapolis, Minnesota: Association for Computational Linguistics.

Naseem, U.; Dunn, A. G.; Kim, J.; and Khushi, M. 2022. Early Identification of Depression Severity Levels on Reddit Using Ordinal Classification. In *WWW*, 2563–2572. ACM.

Organization, W. H.; et al. 2019. Suicide in the world: global health estimates. Technical report, World Health Organization.

Owen, D.; Camacho-Collados, J.; and Anke, L. E. 2020. Towards Preemptive Detection of Depression and Anxiety in Twitter. In *Proceedings of the Fifth Social Media Mining for Health Applications Workshop & Shared Task*, 82–89.

Park, M.; McDonald, D.; and Cha, M. 2021. Perception Differences between the Depressed and Non-Depressed Users in Twitter. *Proceedings of the International AAAI Conference on Web and Social Media*, 7(1): 476–485.

Parraga-Alava, J.; Caicedo, R. A.; Gómez, J. M.; and Inostroza-Ponta, M. 2019. An unsupervised learning approach for automatically to categorize potential suicide messages in social media. In *2019 38th International Conference of the Chilean Computer Science Society (SCCC)*, 1–8. IEEE.

Pennington, J.; Socher, R.; and Manning, C. 2014. GloVe: Global Vectors for Word Representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543. Doha, Qatar: Association for Computational Linguistics.

Posner, K.; Brent, D.; Lucas, C.; Gould, M.; Stanley, B.; Brown, G.; Fisher, P.; Zelazny, J.; Burke, A.; Oquendo, M.; et al. 2008. COLUMBIA-SUICIDE SEVERITY RATING SCALE (C-SSRS).

Rawat, B. P. S.; Kovaly, S.; Yu, H.; and Pigeon, W. 2022. ScAN: Suicide Attempt and Ideation Events Dataset. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1029–1040. Seattle, United States: Association for Computational Linguistics.

Sawhney, R.; Agarwal, S.; Neerkaje, A. T.; Aletras, N.; Nakov, P.; and Flek, L. 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*, 1716–1727.

Sawhney, R.; Joshi, H.; Gandhi, S.; and Shah, R. R. 2021a. Towards ordinal suicide ideation detection on social media. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, 22–30. New York, NY, USA: Association for Computing Machinery.

Sawhney, R.; Joshi, H.; Shah, R. R.; and Flek, L. 2021b. Suicide Ideation Detection via Social and Temporal User Representations using Hyperbolic Learning. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2176–2190. Online: Association for Computational Linguistics.

Sawhney, R.; Manchanda, P.; Singh, R.; and Aggarwal, S. 2018. A Computational Approach to Feature Extraction for Identification of Suicidal Ideation in Tweets. In *Proceedings of ACL 2018, Student Research Workshop*, 91–98. Melbourne, Australia: Association for Computational Linguistics.

Sawhney, R.; Neerkaje, A. T.; and Gaur, M. 2022. A risk-averse mechanism for suicidality assessment on social media. *Association for Computational Linguistics 2022 (ACL 2022)*.

Sher, L. 2020. The impact of the COVID-19 pandemic on suicide rates. *QJM: An International Journal of Medicine*, 113(10): 707–712.

Shing, H.-C.; Nair, S.; Zirikly, A.; Friedenberg, M.; Daumé III, H.; and Resnik, P. 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*, 25–36. New Orleans, LA: Association for Computational Linguistics.

Snell, J.; Swersky, K.; and Zemel, R. 2017. Prototypical Networks for Few-shot Learning. In Guyon, I.; Luxburg, U. V.; Bengio, S.; Wallach, H.; Fergus, R.; Vishwanathan, S.; and Garnett, R., eds., *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

Torjesen, I. 2020. Covid-19: Mental health services must be boosted to deal with "tsunami" of cases after lockdown.

Walsh M., F. P. 1988. Nursing Rituals: Research and Rational Actions. Technical report, Heinemann, London.

Zirikly, A.; Resnik, P.; Uzuner, O.; and Hollingshead, K. 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*, 24–33. Minneapolis, Minnesota: Association for Computational Linguistics.

Zogan, H.; Razzak, I.; Jameel, S.; and Xu, G. 2021. DepressionNet: Learning Multi-modalities with User Post Summarization for Depression Detection on Social Media. In *SIGIR*, 133–142. New York, NY, USA: ACM.

Zogan, H.; Razzak, I.; Wang, X.; Jameel, S.; and Xu, G. 2022. Explainable depression detection with multi-aspect features using a hybrid deep learning model on social media. *World Wide Web*, 25(1): 281–304.

## Paper Checklist

1. Paper Checklist
   (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures?
   Yes, we develop an algorithm integrated with psychology to advance the fusion of social media computation and psychology.
   (b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope?
   Yes, our main claims in the abstract and introduction accurately reflect the paper's contributions and scope.
   (c) Do you clarify how the proposed methodological approach is appropriate for the claims made?
   Yes. The main claims are substantiated by the incorporation of psychological scales, a methodology designed to enhance both performance and interpretability. These scales have undergone validation, bolstering the credibility and efficacy of our approach.
   (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions?
   Yes, we address this in the section on Broader Impact and Ethical Considerations.
   (e) Did you describe the limitations of your work?
   Yes, we address this in the section on Broader Impact and Ethical Considerations.
   (f) Did you discuss any potential negative societal impacts of your work?
   Yes, we address this in the section on Broader Impact and Ethical Considerations.
   (g) Did you discuss any potential misuse of your work?
   Yes, we address this in the section on Broader Impact and Ethical Considerations.
   (h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings?
   Yes, we address this in the section on Broader Impact and Ethical Considerations.
   (i) Have you read the ethics review guidelines and ensured that your paper conforms to them? Yes, we have read the ethics review guidelines.
2. Additionally, if your study involves hypotheses testing...
   (a) Did you clearly state the assumptions underlying all theoretical results? Yes, we have clearly state the assumptions.
   (b) Have you provided justifications for all theoretical results?
   Yes, this study is based on existing findings in psychology.
   (c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results?
   Not applicable
   (d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study?
   Not applicable
   (e) Did you address potential biases or limitations in your theoretical framework?
   Not applicable
   (f) Have you related your theoretical results to the existing literature in social science?
   Yes, this study is based on existing findings in psychology.
   (g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain?
   Yes
3. Additionally, if you are including theoretical proofs...
   (a) Did you state the full set of assumptions of all theoretical results?
   Not applicable
   (b) Did you include complete proofs of all theoretical results?
   Not applicable
4. Additionally, if you ran machine learning experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? Yes, we are willing to provide the code upon identity verification to prevent the misuse.
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? Yes
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? Yes
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? Yes
   (e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? Yes
   (f) Do you discuss what is "the cost“ of misclassification and fault (in)tolerance? Yes
5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets, **without compromising anonymity**...

(a) If your work uses existing assets, did you cite the creators?
Yes

(b) Did you mention the license of the assets?
Yes

(c) Did you include any new assets in the supplemental material or as a URL?
Not applicable

(d) Did you discuss whether and how consent was obtained from people whose data you're using/curating?
Not applicable

(e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content?
Not applicable

(f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR (see **?**)?
Not applicable

(g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset (see **?**)?
Not applicable