Temporal Tides of Emotional Resonance: A Novel Approach to Identify Mental Health on Social Media

Usman Naseem[1], Surendrabikram Thapa[2], Qi Zhang[3,4], Junaid Rashid[5], Liang Hu[3,4], Mehwish Nasim[6,7]
[1]James Cook University, Australia [2]Virginia Tech, USA
[3]Tongji University, China [4]DeepBlue Academy of Sciences, China
[5]Sejong University, South Korea
[6]University of Western Australia [7]Flinders University, Australia

Abstract

Identifying mental health conditions using user-generated text on social media paved the way for automated computational methods for mental health surveillance on social media. Inferring the accurate state of mental health markers requires understanding the user's emotions and history of mental health conditions for identifying the mental health landscape of at-risk users. Existing works have exploited social media data by using users' most recent posts; however, previous works ignored users' emotional historical activity on social media, indicating their mental state over time. In this work, we address this gap and present a novel emotion- and time-aware architecture that jointly learns social media users' emotional historical context and temporal posting irregularities for mental health surveillance on social media. We conduct classification experiments to identify social media users with mental conditions, i.e., (i) healthy users and users affected with depression and (ii) healthy users and users prone to self-harm. Experimental results demonstrate proposed method outperforms recent competitive methods, demonstrating the importance of capturing the user's temporal emotional spectrum and time-aware emotional context through historical posts for social media users' mental health surveillance.

1 Introduction

According to the World Health Organization (WHO), one in four Americans has encountered mental conditions at some point in their lives (Consortium et al., 2004). About 4 – 8% of people in England will experience depression in their lifetime. Mental health has become a global concern and has procured more attention after the COVID-19 pandemic. Due to the high cost of mental health burden, computational research has even focused on analyzing mental health conditions using non-clinical data (Bucci et al., 2019; Moorhead et al., 2013; Adhikari et al., 2022; Thapa et al., 2020). Recent studies have shown that it is possible to learn about social media users' mental health conditions using user-generated text and behavioral analyses of their social media activity (Burke et al., 2010; Naseem et al., 2022a).

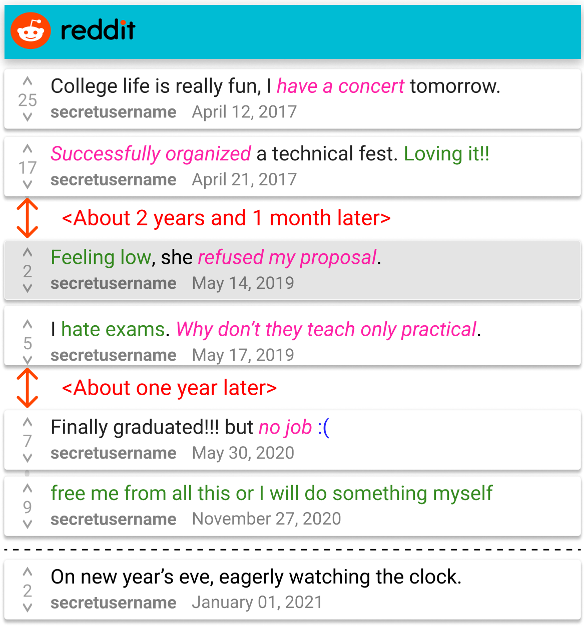

Figure 1: An example of a social media user with a mental health condition. Without analyzing the user's historical posts, which reflect depression inclinations, it is challenging to determine depression. Sequentially evaluating a user's postings without considering timing inconsistencies may misrepresent a user's mental state. Posts are paraphrased for ethical consideration.

In particular, advancement in natural language processing (NLP) techniques and the availability of longitudinal data can play an essential role in assessing user-generated text on social media to identify risk markers for social media users (Naseem et al., 2022e,c). NLP methods can also help create a semi-automated system that can speed up the diagnosis process (Naseem et al., 2022d; Thapa et al., 2022; Adhikari et al., 2021). They may also

Proceedings of the 11th International Workshop on Natural Language Processing for Social Media, pages 1–8
November 1, 2023.

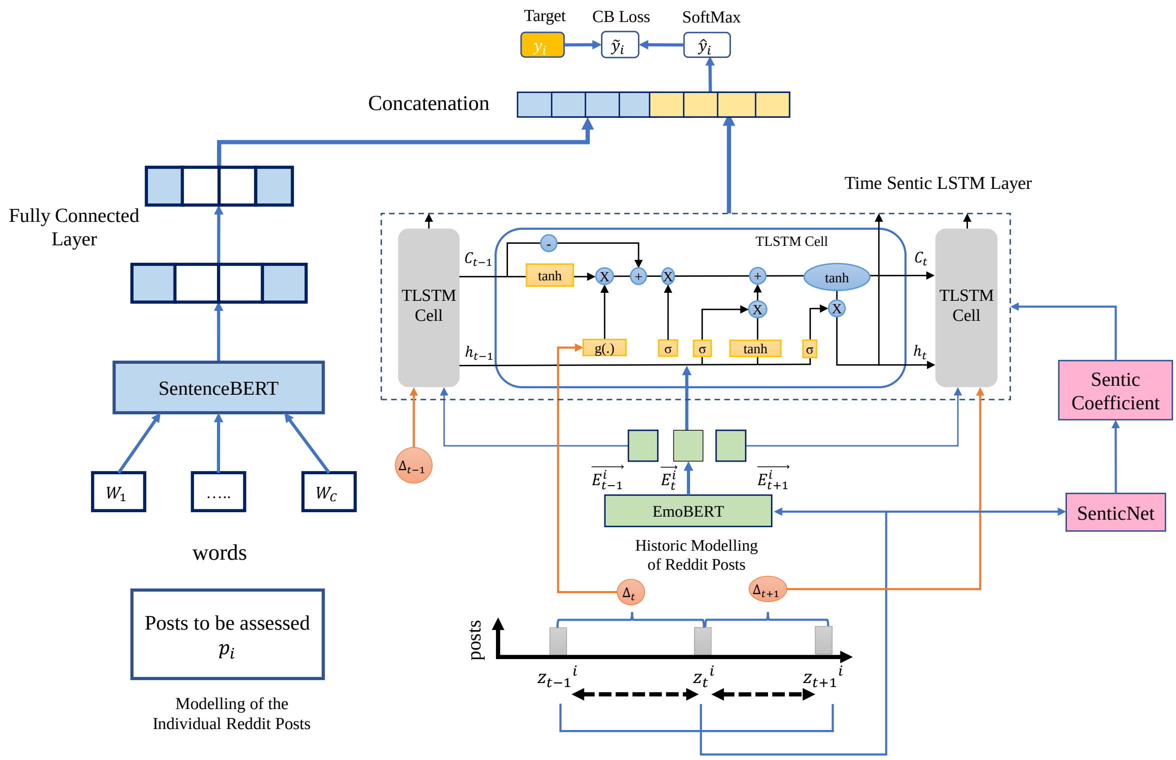

Figure 2: Overall architecture of proposed method

outperform traditional clinical prediction methods to detect mental health issues automatically by improving the specificity and speed of diagnosis (Ríssola et al., 2021). Furthermore, additional features, such as a user's social network (Zhou et al., 2015) or historical posts (Zogan et al., 2021; Naseem et al., 2023, 2022b), may also provide an auxiliary context in recognizing the development of negative emotions that are frequently associated with mental health conditions. Such data is likely to improve the performance of NLP models.

Relevant studies in this area focussed on context-aware sequential modeling (Tsakalidis et al., 2022; Ragheb et al., 2019), one-class-classification approach (Aguilera et al., 2021), time-aware LSTM models (Kang et al., 2022), emotion-based implicit lexicons (Aragón et al., 2019; Barros et al., 2021), and transformer-based classifiers (Martínez-Castaño et al., 2021; Maupomé et al., 2021). These NLP-based studies achieved desirable performance, albeit none of the studies leveraged user context for further performance improvement. We posit that studying users' historical, social media posts and emotion spectrum can help assess their risk of mental health conditions (Figure 1).

Researchers have captured user context as a bag-of-posts (Gaur et al., 2019) or sequentially (Cao et al., 2019; Matero et al., 2019; Zogan et al., 2021) for detecting social media users' mental health conditions. Though such methods are plausible, one of the major shortcomings is the inconsistent time interval between users' historical posts, which can have a critical impact on the correctness of the analysis. This implies that it is important to capture the gap between the user's current posts that are indicative of mental health conditions as well as those from previous years (Figure 1). Such temporal irregularities in users' historical posts affect users' evaluations. Sequential methods such as Long-Short-Term-Memory (LSTM) networks assume regular posting intervals, limiting the learning ability of a user's emotion spectrum over varied time intervals. In this study, we address these challenges. Our main contributions include the following:

- We present a novel architecture that jointly learns a user's emotional historical context and temporal posting irregularities for early detection of mental health issues on social media (section 2);
- We introduce Time-Sentic-LSTM, an extension of a time-aware LSTM, by incorporating components accounting for emo-

tional concept-level user's representation (section 2.2);

- Experimental results demonstrate that our method outperforms previous state-of-the-art methods for early detection of mental health issues using social media data (section 4).

2 Method

Overview of our architecture: The architecture of the proposed method is shown in Figure 2, which consists of the current post representation layer and the representation of the user's emotional historical spectrum, where we first encoded the individual historical posts and then modeled historical posts sequentially. In the subsequent discussion, we will explain each module in depth.

Problem Definition: We formally define the problem: given a social media user $u_i \in \{u_1, u_2, ..., u_n\}$ with historical posts $p_i \in \{p_1, p_2, ..., p_n\}$, *i* represents the number of posts. Our objective is early risk identification of the user with 1) depression and 2) self-harm risk using a user-level historical timeline. Each user post p_i is associated with history $Z_{i,j} = (z_1^i, \gamma_1^i), (z_2^i, \gamma_2^i), \cdots, (z_L^i, \gamma_L^i)$ where z_t^i is a historic post by the user u_j posted at time γ_k^i with $\gamma_1^i < \gamma_2^i <, \cdot, < \gamma_L^i < \gamma_{current}^i$. We formulate our problem as a binary classification task to predict a label y_i to the user u_i from the corresponding set of labels.

2.1 Current Post Representation

We used SentenceBERT (Reimers and Gurevych, 2019) to capture the linguistic features of a current post P_i. SentenceBERT calculates the mean of word output vectors to generate a fixed-size sentence embedding. This is formulated in eq. (1).

$$P_i = SentenceBERT(p_i) \tag{1}$$

where $P_i \in \mathbf{R}^{768}$ is transformed linearly using a dense layer with 768-dimensions.

2.2 User's Emotional Historic Spectrum

Individual historic post representation: Escalation in emotional aspects may represent an increased risk of depression and self-harm in Reddit posts. Hence, we extract the emotional spectrum of each historic post (z_k^i). General text encoders fail to capture fine-grained emotions from users' historical posts on social media. To mitigate that shortcoming, we capture fine-grained emotions using EmoBERT (Aduragba et al., 2021), a fine-tuned variant of BERT over fine-grained emotions of social media data associated with social well-being and reflected in implicit domain-specific emotions in posts.

Using EmoBERT, we tokenized each historical post and added the `[CLS]` token at the beginning of each post and used the final hidden state corresponding to the `[CLS]` token of 768-dimension as the aggregate representation of the emotional spectrum. We defined our emotional vector (E_k^i) $\in \mathbf{R}^{768}$ of each historic post z_k^i:

$$E_k^i = EmoBERT(z_k^i) \tag{2}$$

Emotional concept-level representation of user's historical posts: We introduced the use of SenticNet (Cambria et al., 2018) to extract the emotional concept-level representation of the user's historical posts and fuse it into the end-to-end training of an extension of an LSTM cell. To achieve this, a set of *C* concept candidates would be extracted using a syntactic concept parser and mapped to the d_c dimensional vectors $[\alpha_{t,1}, \alpha_{t,2}, ..., \alpha_{t,C}]$ at time step *t*. The candidate embedding of step *t* is calculated as the average of the vectors using eq 3:

$$\alpha_t = \frac{1}{C} \sum_i \alpha_{t,i} \tag{3}$$

Our extension of an LSTM cell is formulated as:

$$\begin{aligned}
f_t &= \sigma(W_f[x_t, h_{t-1}, \alpha_t] + b_f) \\
I_t &= \sigma(W_I[x_t, h_{t-1}, \alpha_t] + b_I) \\
\hat{C}_t &= \tanh(W_C[x_t, h_{t-1}] + b_c) \\
C_t &= f_t * C_{t-1} + I_t * \hat{C}_t \\
o_t &= \sigma(W_o[x_t, h_{t-1}, \alpha_t] + b_o \\
o_t^c &= \sigma(W_{co}[x_t, h_{t-1}, \alpha_t] + b_{co} \\
h_t &= 0_t * \tanh(C_t) + 0_t^c * \tanh(W_c \alpha_t)
\end{aligned}$$

where f_t, I_t, and o_t are the forget gate, input gate, and output gate, respectively. W_f, W_I, W_o, b_f, b_I and b_o are the weight matrix and bias scalar for each gate. C_t is the cell state, and h_t is the hidden output. The extracted emotional concepts are added to the forget, input, and output gates of an extended LSTM cell to capture sequential modeling of the user's historical posts.

Sequential historic post modeling: To capture the users' historical posts with irregular time intervals sequentially, we propose leveraging Time-aware LSTM (T-LSTM) (Baytas et al., 2017) where time lapse between posts is forwarded to the T-LSTM cell (Figure 2). The T-LSTM cell combines the actual time interval between posts, the emotional concept-level representation of the user's historical posts, and the emotional context of each historical post (E_k^i).

T-LSTM weights the short-term memory cell (C_t^S) and introduces memory time decay dependent on the time between consecutive posts. T-LSTM leverages a monotonically declining elapsed time mechanism to transform time into appropriate weights to achieve less impact on posts with larger elapsed time between posts. T-LSTM incorporates time lapses as follows:

$$C_{t-1}^S = \tanh(W_d C_{t-1} + b_d) \quad \text{(Short-term memory)}$$
$$\hat{C}_{t-1}^S = C_{t-1}^S * g(\Delta_t) \quad \text{(Discounted short-term memory)}$$
$$C_{t-1}^T = C_{t-1} + \hat{C}_{t-1}^S \quad \text{(Long-term memory)}$$
$$C_{t-1}^* = C_{t-1}^T + \hat{C}_{t-1}^S \quad \text{(Adjusted previous memory)}$$

where c_{t-1} and c_t are previous and current cell memories, and $\{W_d, b_d\}$ are model parameters. Δ_t is the time between user's historic posts i.e., z_{t-1} and z_t, $g(\Delta_t)$ is the decaying mechanism used in by Baytas et al. (2017). To derive the current hidden state ($\hat{H}_t^i$), T-LSTM changes LSTM gate operations by incorporating C_{t-1}^* in place of C_t.

2.3 Joint Network Optimization

The proposed method jointly learns the language of the current post and the user's emotional historic spectrum in a time-aware manner. To achieve this, we concatenate the encoded features of a current post (P_i) and $\hat{H}_t^i$ followed by a softmax function over a dense layer with Rectified Linear Unit (ReLU) to obtain the probability of the early risk detection. This is formulated as follows:

$$\begin{aligned} y_i = RELU(W_y(P_i \oplus \hat{H}_t^i) + b_y) \\ \hat{y}_i = softmax(y_i) \end{aligned} \tag{4}$$

where $\hat{y}_i$ is the final output and W_y, b_y are the model parameters.

To address the class imbalance, we trained our model using class balanced loss (Cui et al., 2019) along with focal loss (Lin et al., 2017). This loss method assigns class-wise re-weighting by incorporating an inversely proportionate weighting component and is defined as:

$$L = CB_{focal}(\hat{y}, y_i; \beta, \gamma) \tag{5}$$

where $\hat{y}_i$ is the predicted label, CB_{focal} is the class-balanced focal loss, and y_i is the label of the current tweet. β and γ are the hyperparameters.

3 Experimental Setup

3.1 Datasets

We evaluated the performance of our method on two datasets from CLEF eRISK[1] challenges: early risk detection of (i) self-harm and (ii) depression. We combined and used the data from the last three years in both datasets. The self-harm dataset includes 2,068 users labeled as *self-harm* (1,785) and *non-self-harm* (283). The depression dataset consists of 170 users labeled as *depressed* (16 users) and *non-depressed* (154 users). The statistics of the datasets we used are given in Table 1.

Stats\Datasets	Self-harm	Depression
No. of users	2,068	170
Class-wise distribution		
No self-harm/depression	44.58%	9.40%
self-harm/depression	19.53%	90.60%

Table 1: Datasets Statistics

3.2 Experimental Settings

For consistency, we used the same experimental settings for all models and used 10-fold cross-validation. All results are reported as the average across all folds. We used the grid search optimization technique to optimize the parameters. To tune the number of layers (n), we empirically experimented with the values: n ∈ {1, 2, 3}. Similarly, for drop out (δ), we tested with: δ ∈ {0, 0.2, 0.4, 0.5, 0.8}, hidden dimension (H) with H ∈ {64, 128, 256}, learning rate (lr): lr ∈ {0.001, 0.005, 0.01, 0.02} and control parameter β ∈ {0, 0.3, 0.6,....,3.0}. For optimization (O): O ∈ {'Adam', 'Adamax', 'AdamaW'} with a batch size of 16 were used. We used base version pre-trained language models (LMs) using HuggingFace[2], an open-source Python library. Varying lengths of posts are padded and trained for 150 epochs with early stopping with a patience of 10 epochs. The hyperparameters used in our experiments are n = 2, δ = 0.5, H = 128, O = AdamW, lr = 0.005, and β = 2. We will release our code publicly available for reproduction once the review is finished.

[1] https://erisk.irlab.org/
[2] https://huggingface.co/models

Baselines Methods\Datasets		Depression			Self Harm		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score
Post-level	RF	0.45	0.50	0.47	0.44	0.50	0.47
	GRU	0.60	0.63	0.61	0.59	0.58	0.58
	LSTM	0.52	0.51	0.51	0.59	0.59	0.59
	C-LSTM	0.42	0.50	0.46	0.55	0.58	0.56
User-level	Contextual CNN	0.42	0.50	0.46	0.50	0.49	0.49
	Suicide Detection Model	0.46	0.47	0.46	0.57	0.58	0.57
	DualContextBert	0.56	0.63	0.59	0.62	0.56	0.59
	STATENet	0.57	0.56	0.56	0.56	0.54	0.55
	DepressionNet	0.61	0.63	0.62	0.61	0.59	0.60
Our method		**0.72****	**0.67****	**0.69****	**0.63****	**0.62****	**0.62****

Table 2: Comparison of the proposed method v/s the baselines. F1, Precision, and Recall scores are averaged over 10 folds. * indicates that the proposed method achieved a significant ($p < 0.05$) performance improvement over the best baseline under Mann–Whitney U test.

3.3 Baselines

We validate the effectiveness of our model by comparing it with *post-level* and *user-level* baseline methods, which are widely used in previous similar studies.

- **Post-level baselines**

  Random Forest (RF): A non-contextual PL approach that applies random forest on their statistical and linguistic features (Sawhney et al., 2018b).

  Gated Recurrent Unit (GRU): We apply GRU model which undergo faster training of training data as compared to that of LSTM due to less number of parameters (Cho et al., 2014).

  Long-Short Term Memory (LSTM): We apply LSTM model for mapping emotional spectrum from historical timeline.

  Contextualized LSTM (C-LSTM): C-LSTM deep neural network uses CNN for feature extraction and LSTM for encoding posts (Sawhney et al., 2018a).

- **User-level baselines**

  Contextual CNN: A non-sequential UL model over posts encoded with GloVe embedding (Gaur et al., 2019).

  Suicide Detection Model: A user-level attention based LSTM model that encodes posts using FastText embedding (Cao et al., 2019).

  DualContextBERT: The best performing model at CLPsych 2019 which feeds BERT encoded posts to the attention-based RNN layer (Matero et al., 2019).

  STATENet: Suicidality assessment Time-Aware TEmporal Network, a transformer-based framework to identify suicidal risk on social media. STATENet uses a dual transformer-based architecture to learn tweets' linguistic and emotional cues (Sawhney et al., 2020).

  DepressionNet: A novel approach which summarizes user posts before encoding it via embedding. Authors apply the BiGRU model and concatenate the results with encoded current post (Zogan et al., 2021).

4 Results

4.1 Overall Comparison

Results in Table 2 explain the efficacy of our method as it outperforms the existing baselines on both datasets. The results with F1-Score are 0.69 and 0.62 for the depression and self-harm dataset, respectively, which is an absolute increase of 7% and 2%, respectively, compared to the DepressionNet. We observe better performance with the use of contextual features as compared to non-contextual feature-based methods. We attribute this to the reason that temporal context-based methods capture better insight into the mental state of users. Methods that sequentially capture the user's historical

Model	Depression	Self Harm
Current post only	0.54	0.56
Current + Sequential	0.59	0.59
Current + Time-aware	0.62	0.61
Our method	**0.69***	**0.62***

Table 3: Ablation analysis: F1-Scores are averaged over 10 folds. *indicates that the proposed method achieved a significant ($p < 0.05$) performance improvement over other variants under the Mann–Whitney U test.

post perform better than other methods, such as Context CNN. Recent advancement with DepressionNet recapitulates their ability to model temporal dependencies of historical posts compared to Context CNN's and DepressionNet. Our method performs better than other sequential-based baselines because it captures irregular time intervals in the posting history of a user using time-aware modeling, which enables our model to capture the user's emotional historical context accurately.

4.2 Ablation Study

We conduct an ablation study to evaluate the effectiveness of each component of our method (Table 3). The F1-Score drops on both datasets when we use only the current post of a user. The performance improves with sequential information followed by a time-aware model that captures irregular time intervals of users' historical postings and the user's emotional historic spectrum-based features. We conclude that the strengths of our method lie in the use of current post representation and the user's emotional historical spectrum.

5 Ethical Considerations

Our empirical work on the user's social media timeline does abide by ethical considerations. The metadata of self-reported writing on social media is sensitive and contains personal information and demographics. In this context, we incorporate the trade-off between privacy and the effectiveness of the proposed method. We present our work in a non-intrusive manner and do not specify any real-time examples in this research paper. We used the dataset available in the shared task of the eRISK workshop in the CLEF forum, studied it purely observationally, and did not intervene in the user's personal information. The annotated datasets we used are publicly available and include de-identified publicly available posts where users understand public access and there is no expectation of privacy. Hence, no ethical approval is required for this research.

6 Conclusion

We present a new emotion and time-aware architecture for social media users' mental health surveillance. Our method takes inspiration from psychological studies about analyzing a user's temporal emotional spectrum and capturing the time-aware emotional context of users through historical posts for more accurate mental health surveillance of social media users with self-harm and depression. The experimental results showed that our method outperformed previous methods for monitoring mental health conditions on social media.

Limitations

While our study presents a novel approach to mental health surveillance on social media, several limitations should be acknowledged. The reliance on user-generated text data assumes that individuals openly share their mental health experiences online, potentially introducing selection bias. Emotion detection from text data, while valuable, may not capture the full complexity of emotional states. The model's generalizability to diverse populations, platforms, and cultural contexts may require further validation. Interpretability and fairness in predictions, along with potential long-term impacts on individuals, deserve ongoing scrutiny.

References

Surabhi Adhikari, Surendrabikram Thapa, Usman Naseem, Priyanka Singh, Huan Huo, Gnana Bharathy, and Mukesh Prasad. 2022. Exploiting linguistic information from nepali transcripts for early detection of alzheimer's disease using natural language processing and machine learning techniques. *International Journal of Human-Computer Studies*, 160:102761.

Surabhi Adhikari, Surendrabikram Thapa, Priyanka Singh, Huan Huo, Gnana Bharathy, and Mukesh Prasad. 2021. A comparative study of machine learning and nlp techniques for uses of stop words by patients in diagnosis of alzheimer's disease. In *2021 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE.

Olanrewaju Tahir Aduragba, Jialin Yu, Alexandra I Cristea, and Lei Shi. 2021. Detecting fine-grained emotions on social media during major disease outbreaks: Health and well-being before and during the covid-19 pandemic. In *AMIA Annual Symposium Proceedings*, volume 2021, page 187. American Medical Informatics Association.

Juan Aguilera, Delia Irazú Hernández Farías, Manuel Montes-y Gómez, and Luis C González. 2021. Detecting traces of self-harm in social media: A simple and interpretable approach. In *Mexican International Conference on Artificial Intelligence*, pages 196–207. Springer.

Mario Ezra Aragón, Adrián Pastor López-Monroy, and Manuel Montes-y Gómez. 2019. Inaoe-cimat at erisk 2019: Detecting signs of anorexia using fine-grained emotions. In *CLEF (Working Notes)*.

Lucas Barros, Alina Trifan, and José Luís Oliveira. 2021. Vader meets bert: sentiment analysis for early detection of signs of self-harm through social mining. In *CLEF (Working Notes)*, pages 897–907.

Inci M Baytas, Cao Xiao, Xi Zhang, Fei Wang, Anil K Jain, and Jiayu Zhou. 2017. Patient subtyping via time-aware lstm networks. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 65–74.

Sandra Bucci, Matthias Schwannauer, and Natalie Berry. 2019. The digital revolution and its impact on mental health care. *Psychology and Psychotherapy: Theory, Research and Practice*, 92(2):277–297.

Moira Burke, Cameron Marlow, and Thomas Lento. 2010. Social network activity and social well-being. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1909–1912.

Erik Cambria, Soujanya Poria, Devamanyu Hazarika, and Kenneth Kwok. 2018. Senticnet 5: Discovering conceptual primitives for sentiment analysis by means of context embeddings. In *AAAI*.

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. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 1718–1728.

Kyunghyun Cho, Bart Van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. 2014. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259*.

WHO World Mental Health Survey Consortium et al. 2004. Prevalence, severity, and unmet need for treatment of mental disorders in the world health organization world mental health surveys. *Jama*, 291(21):2581–2590.

Yin Cui, Menglin Jia, Tsung-Yi Lin, Yang Song, and Serge Belongie. 2019. Class-balanced loss based on effective number of samples. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9268–9277.

Manas Gaur, Amanuel Alambo, Joy Prakash Sain, Ugur Kursuncu, Krishnaprasad Thirunarayan, Ramakanth Kavuluru, Amit Sheth, Randy Welton, and Jyotishman Pathak. 2019. Knowledge-aware assessment of severity of suicide risk for early intervention. In *The World Wide Web Conference*, pages 514–525.

Xin Kang, Rongyu Dou, and Haitao Yu. 2022. Tua1 at erisk 2022: Exploring affective memories for early detection of depression.

Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. 2017. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988.

Rodrigo Martínez-Castaño, Amal Htait, Leif Azzopardi, and Yashar Moshfeghi. 2021. Bert-based transformers for early detection of mental health illnesses. In *International Conference of the Cross-Language Evaluation Forum for European Languages*, pages 189–200. Springer.

Matthew Matero, Akash Idnani, Youngseo Son, Salvatore Giorgi, Huy Vu, Mohammad Zamani, Parth Limbachiya, Sharath Chandra Guntuku, and H Andrew Schwartz. 2019. Suicide risk assessment with multi-level dual-context language and bert. In *Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology*, pages 39–44.

Diego Maupomé, Maxime D Armstrong, Fanny Rancourt, Thomas Soulas, and Marie-Jean Meurs. 2021. Early detection of signs of pathological gambling, self-harm and depression through topic extraction and neural networks. In *CLEF (Working Notes)*, pages 1031–1045.

S Anne Moorhead, Diane E Hazlett, Laura Harrison, Jennifer K Carroll, Anthea Irwin, and Ciska Hoving. 2013. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. *Journal of medical Internet research*, 15(4):e1933.

Usman Naseem, Adam G Dunn, Jinman Kim, and Matloob Khushi. 2022a. Early identification of depression severity levels on reddit using ordinal classification. In *Proceedings of the ACM Web Conference 2022*, pages 2563–2572.

Usman Naseem, Matloob Khushi, Jinman Kim, and Adam G Dunn. 2022b. Hybrid text representation for explainable suicide risk identification on social media. *IEEE transactions on computational social systems*.

Usman Naseem, Jinman Kim, Matloob Khushi, and Adam Dunn. 2023. Graph-based hierarchical attention network for suicide risk detection on social media. In *Companion Proceedings of the ACM Web Conference 2023*, pages 995–1003.

Usman Naseem, Jinman Kim, Matloob Khushi, and Adam G Dunn. 2022c. Identification of disease or symptom terms in reddit to improve health mention classification. In *Proceedings of the ACM Web Conference 2022*, pages 2573–2581.

Usman Naseem, Jinman Kim, Matloob Khushi, and Adam G Dunn. 2022d. Robust identification of figurative language in personal health mentions on twitter. *IEEE Transactions on Artificial Intelligence*, 4(2):362–372.

Usman Naseem, Byoung Chan Lee, Matloob Khushi, Jinman Kim, and Adam G Dunn. 2022e. Benchmarking for public health surveillance tasks on social media with a domain-specific pretrained language model. *NLP-Power 2022*, page 22.

Waleed Ragheb, Jérôme Azé, Sandra Bringay, and Maximilien Servajean. 2019. Attentive multi-stage learning for early risk detection of signs of anorexia and self-harm on social media. In *CLEF 2019 Working Notes-Conference and Labs of the Evaluation Forum*, volume 2380, page 126.

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

Esteban A Ríssola, David E Losada, and Fabio Crestani. 2021. A survey of computational methods for online mental state assessment on social media. *ACM Transactions on Computing for Healthcare*, 2(2):1–31.

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.

Ramit Sawhney, Prachi Manchanda, Puneet Mathur, Rajiv Shah, and Raj Singh. 2018a. Exploring and learning suicidal ideation connotations on social media with deep learning. In *Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 167–175.

Ramit Sawhney, Prachi Manchanda, Raj Singh, and Swati Aggarwal. 2018b. A computational approach to feature extraction for identification of suicidal ideation in tweets. In *Proceedings of ACL 2018, Student Research Workshop*, pages 91–98.

Surendrabikram Thapa, Surabhi Adhikari, Usman Naseem, Priyanka Singh, Gnana Bharathy, and Mukesh Prasad. 2020. Detecting alzheimer's disease by exploiting linguistic information from nepali transcript. In *Neural Information Processing: 27th International Conference, ICONIP 2020, Bangkok, Thailand, November 18–22, 2020, Proceedings, Part IV 27*, pages 176–184. Springer.

Surendrabikram Thapa, Awishkar Ghimire, Surabhi Adhikari, Akash Kumar Bhoi, and Paolo Barsocchi. 2022. Cognitive internet of things (iot) and computational intelligence for mental well-being. In *Cognitive and Soft Computing Techniques for the Analysis of Healthcare Data*, pages 59–77. Elsevier.

Adam Tsakalidis, Federico Nanni, Anthony Hills, Jenny Chim, Jiayu Song, and Maria Liakata. 2022. Identifying moments of change from longitudinal user text. *arXiv preprint arXiv:2205.05593*.

Xujuan Zhou, Enrico Coiera, Guy Tsafnat, Diana Arachi, Mei-Sing Ong, Adam G Dunn, et al. 2015. Using social connection information to improve opinion mining: Identifying negative sentiment about hpv vaccines on twitter.

Hamad Zogan, Imran Razzak, Shoaib Jameel, and Guandong Xu. 2021. Depressionnet: A novel summarization boosted deep framework for depression detection on social media. *arXiv preprint arXiv:2105.10878*.