Code RED: Reactive Emotion Difference for Stress Detection on Social Media.

Zeyad Alghamdi, Faisal Alatawi, Mansooreh Karami, Tharindu Kumarage, Ahmadreza Mosallanezhad, and Huan Liu

School of Computing and Augmented Intelligence
Arizona State University, Tempe, USA
{zalgham1,faalataw,mkarami,kskumara,amosalla,huanliu}@asu.edu

Abstract. The prevalence of stress-related issues has become increasingly evident in recent years. Many of the population use social media platforms as an outlet to talk about life situations and express stress. Furthermore, detecting stress has become a critical matter due to the impact of stress on our daily lives, and the consequences of not detecting stress early on can lead to severe physical and mental health complications. Therefore, social media stress detection is an emerging field that leverages machine learning and deep learning techniques to identify stress indicators in social media posts. While most of the works in this field focus on analyzing the posts' textual contents, many ignore the social support cues that could aid the stress detection process. This study proposes a stress detection method by leveraging the emotional content of posts, social support or comments inspired by multidisciplinary (social sciences) theories. We build a classifier based on the emotional difference between the initial post and the responses or reactions it receives. We utilize a state-of-the-art transformer-based emotion classifier and two publicly available datasets. Our approach achieves a better stress classification by incorporating social support emotions. Our main contributions are the novelty and utility of the new approach to stress detection on social media and the expansion of the datasets to include social support. This study showcases and proves the validity of using social support to detect stress on social media.

Keywords: Social Media · Stress Detection · Emotion Analysis · Social Support · Mental Health · Natural Language Processing

1 Introduction

In recent years, the COVID-19 pandemic has significantly impacted the mental health and well-being of the global population. This impact has catalyzed an approximately 25% increase in mental health condition rates compared to the pre-pandemic era. As recently reported by the World Health Organization (WHO), one in eight individuals is living with a mental health condition, equating to approximately one billion people worldwide[1]. In the United States specifically,

[1] www.who.int/publications/i/item/9789240049338

approximately 21% of adults are reportedly experiencing a mental illness [2]. Stress is hypothesized to pave the way for various adverse health outcomes, culminating not only in severe physical health conditions [12, 21] but also fostering a climate for mental health issues such as anxiety[3] [6], depression [3], and addiction [7, 16]. Despite increased awareness and advocacy efforts, mental health problems remain stigmatized, with societal attitudes still proving challenging. However, the anonymity offered by social media platforms like Reddit has empowered users to discuss their mental health struggles openly. These platforms have become vital outlets for individuals to articulate their emotions, connect with others facing similar issues, and access valuable mental health resources, including online support groups [15, 17]. As a result, social media data now represent a valuable asset for studying and addressing mental health-related challenges. Therefore, researchers were encouraged to analyze social media data for stress detection and other mental health conditions. For this means, they have used different methods for analyzing stress in textual data such as rule-based method [18], Latent Dirichlet Allocation (LDA) [8, 13], and pre-trained language models [14]. Moreover, some also utilized other modalities besides text [10, 11], such as images and social network information. However, we argue that social support (i.e., the community's response to stressed users) can be utilized as auxiliary information to further improve the process of stress detection on social media. For example, researchers show that the emotions expressed by social support convey more positive expressions than those expressed by the stressed user [4]. This is mainly because supportive interactions include more positive feedback, as illustrated in the example in Figure 1, such as appraisal, self-esteem, and belonging support [2, 9]. In contrast, stressed users express more negative emotions, such as sadness, fear, and anger [1, 20].

To this means, we hypothesize that incorporating this variation between the expressed emotions of the poster and commenter as a feature can enhance stress detection capabilities. Specifically, we focus on the individual emotional differences between the original post and each corresponding comment, leading us to our proposed model: Reactive Emotions Difference (RED).

Our contributions in this work are as follows:

- Proposing a novel social media stress detection approach considering the emotional divergence between posters and their social media supporters,
- Augmenting the existing real-world datasets to include social support by extracting the comments for labeled stress and non-stress posts, and
- Providing insights into the effectiveness of responsive emotions difference through correlation analysis.

2 Related work

Researchers have extensively utilized various elements of social media to study stress, with a primary focus on textual posts. Thelwall, for instance, has experi-

[2] https://mhanational.org/issues/state-mental-health-america
[3] www.nimh.nih.gov/health/publications/so-stressed-out-fact-sheet

mented with stress classifiers in social media posts, employing dictionaries, lexicons, and corpus for analysis [18]. There have also been efforts to infer stress from social media posts through lexical approaches and a set of rules, which provided practical applications, albeit less accurate than machine learning models. LDA has been used to create distributions of textual documents and words by topic [8, 13]. A detailed study was also conducted on finer-grained features, employing lexical features, the Dictionary of Affect in Language (DAL), a comprehensive suite of LIWC features, and sentiment through the Pattern sentiment library, along with several syntactic features [19]. Other researchers used multiple input representations and various neural and non-neural models for classification. In both instances, pretraining and fine-tuning a BERT-based language model achieved the best performance [14].

Conversely, some researchers concentrated on user activity and social interactions, incorporating additional features to aid stress classification. Lin et al., for example, aimed to develop a psychological profile and individual characteristics, utilizing photos collected from Facebook profiles as input data. They used Spearman's correlation coefficient analysis and logistic regression as their analysis method and classifier, respectively [10]. They also incorporated statistical data, such as the number of messages from the user and the number of responses, collected over a specific period. Another approach based on tweet text and social relationships was proposed by Lin et al., in which they implemented a hybrid model consisting of a factor graph model and a convolutional neural network (CNN) to enhance detection performance [11].

Previous studies primarily focused on lexical and sentiment features, with a few like Alghamdi et al. [1] and Turcan et al. [20], exploring the role of emotions in stress detection. These studies confirmed the potential of emotions as features to enhance stress detection. However, none have previously incorporated social support emotions into the social media stress detection process.

Zhang et al [22] have explored the use of comment emotions for classification tasks, extracting emotion-guided features from posts and comments, and applying deep neural networks to understand their interplay. Although the study focused on fake news detection, its methodology, based on the premise that emotionally arousing content garners attention, resonates with our approach of integrating social emotions as key features, underscoring the importance of comment analysis in social media studies.

3 Proposed Model

This section provides an overview of the proposed experimental model, which comprises three primary components: text embedding, emotion extraction, and classification, as depicted in Figure 1. Each component will be discussed in detail in the following subsections. It's important to note that the model inputs are the posts and their corresponding comments, and the output is the stress label.

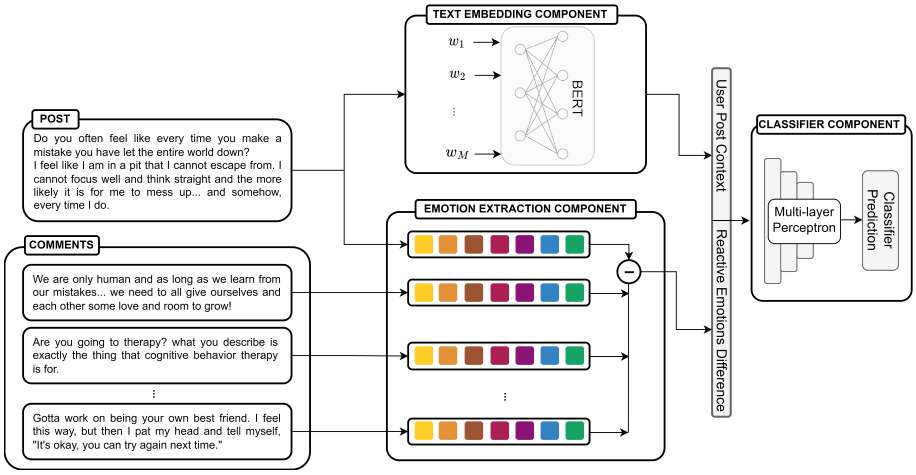

Fig. 1. The model overview shows the three main components: text embedding, emotion extraction (the colors represent different emotions), classifier components, and an example post with a sample of its corresponding comments.

3.1 Text Embedding Component

The initial stage of detecting stress on social media is centered on the extraction of essential textual content from user-generated posts. Such content typically provides insight into their current state of affairs, emotions, or generalized expressions of distress. In order to achieve this, we harness the capabilities of Sentence BERT (SBERT)[4], a tool renowned for its efficacy to extract post embeddings. To better represent our formulation, we let $\mathcal{P} = \{P_i\}_{i=1}^{N}$ denote the set of all posts, where N is the total number of posts. And each P_i is a unique post in $\mathcal{P}$. Each post is a sequence of words $P_i = (w_1, w_2, ..., w_n)$ and the Sentence BERT embedding for the whole post as :

$$\boldsymbol{Emb}_P = SBERT(w_1, w_2, ..., w_n) \tag{1}$$

A further vital aspect of this process involves distilling the emotions encapsulated within these posts.

3.2 Emotions Extraction Component

To extract the emotions expressed within each of the posts and their corresponding comments, we utilize a state-of-the-art transformer emotion classifier model[5]. This model has been fine-tuned on a diverse range of text types, including Twitter, Reddit, student self-reports, and TV dialogues. The output of the model is a vector of six emotions - anger (e_1), fear (e_2), happiness (e_3), joy (e_4),

[4] https://www.sbert.net/
[5] https://huggingface.co/j-hartmann/emotion-english-distilroberta-base

sadness (e_5), and surprise (e_6), along with a neutral (e_7) category. The emotion categories are based on Ekman's theory of basic emotions [5].

We denote the output vector of the model as:

$$\boldsymbol{Emo} = [e_1, e_2, ..., e_7] \tag{2}$$

where each e_{i} denotes the score for the respective emotion. Moreover,

$$0 \leq e_{\mathrm{i}} \leq 1, \text{ and } \sum_{i=1}^{7} e_i = 1 \tag{3}$$

Upon application of the model to the post and its accompanying comments, we derive two fundamental sets of features: a vector that encapsulates the emotional scores for the post, represented as Emo_p, and a collection of vectors signifying the emotional scores of the respective comments, denoted as $[Emo_{c_1}, Emo_{c_2}, ..., Emo_{c_n}]$. For every post P_i, a corresponding set of comments is also defined, symbolized as $\mathcal{C}_i = \{C_{i,j}\}_{j=1}^{M_i}$, where M_i refers to the quantity of comments tied to the post P_i, and $C_{i,j}$ denotes the j^{th} comment associated with the i^{th} post.

Fundamentally, our goal is to utilize the emotions of the users who wrote the post and the emotions of users who replied to the post. That is where our proposed model, 'RED', which stands for the Reactive Emotions Difference, comes into play. Formally, given a post P with an emotion vector Emo_p and a comment C with an emotion vector Emo_c, the RED is computed as:

$$RED(P, C) = ||Emo_p - Emo_c||_1 \tag{4}$$

where $||.||_1$ represents the L1 norm (the absolute value of the differences) between the emotional categories of the post minus their equivalent emotional categories of the comment. The output of this component is our primary feature.

3.3 Classifier Component

The objective of this component is to take the outputs of the previous parts and concatenate them, then build a multi-layer perceptron (MLP) stress detection classifier, which we call ***Code RED***. We release our implementation code and details on GitHub[6].

We implement six hidden layers MLP to the concatenation of the resulting vectors in *Equation*(1) and *Equation*(5). Mathematically, this can be represented as:

$$\boldsymbol{Code\ RED} = \hat{y} = \mathrm{MLP}([RED_{(P,C)} \oplus Emb_P]) \tag{5}$$

[6] github.com/Zeyad-o/CODE-RED

4 Experiments

In this section, we show the performance of our model in detecting stress in social media. In order to demonstrate our model's ability to detect stress, we designed two research questions:

- **RQ1** - Which elements of the RED contribute to the prediction?
- **RQ2** - Can we use RED to detect stress and how does it compare to the baselines?

4.1 Datasets

As a global hub, Reddit provides a semi-anonymous platform enabling narratives from diverse and elusive populations. Its structure, consisting of user-created subreddits, promotes inclusive discourse and community support. This, coupled with its broad reach, allows for in-depth insights into mental health issues, potentially guiding targeted interventions[17]. Our study utilizes two public social media stress analysis datasets derived from this platform. We use the Reddit API Wrapper (PRAW)[7] to extract all feasible comments from available posts. The characteristics of the compiled datasets, which underpin our experiments, are detailed in Table 1.

Datasets	**Size**	**Collected Comments, Avg per post**	**Label**		**comment** $\overline{length}$	$\overline{comments}$ **per label**
Dreadit [19]	2420	25103	Stress	1103	79.49	9.62
		10.37	Non-Stress	1317	72.6	11.69
Rastogi et al. [14]	1646	43826	Stress	927	53.92	29.05
		26.6	non-Stress	719	25.11	22.86

Table 1. Characteristics of retrieved datasets, the overline represents the average number of the entity.

Dreaddit We utilize the Dreaddit dataset in this study, which comprises 3,553 Reddit post segments collected from various support communities during 2017 and 2018, all of which present a strong propensity for stress expression [19]. To annotate the data, crowd-sourcing annotators labeled each post as stress and not stress post. The final label is the majority label. From the original dataset, we retrieved 2,420 posts.

[7] https://praw.readthedocs.io/en/stable/

Social Media Articles by Rastogi et al. Our study also incorporates a dataset curated by [14], specifically designed for stress detection in social media texts. This dataset includes stress-labeled examples drawn from subreddits associated with negative emotions like stress, anxiety, and depression, while non-stress examples were derived from subreddits expressing positive emotions. To guarantee consistent, high-quality samples, automated denoising and annotation methods were utilized. We chose the dataset with the highest number of posts "Reddit Title", which contains 5556 posts. We retrieved 1646 posts from the original dataset.

4.2 Experimental Settings

The success of any machine learning experiment heavily relies on the careful selection and tuning of hyperparameters. In this section, we outline some of the hyperparameters utilized in the MLP model. However, for the sake of brevity, we encourage readers to check the project GitHub page for more details[8].

In our experiment, the input size is determined by the concatenation of the emotion features of comments and the BERT embeddings, resulting in a comprehensive feature representation. As we recall earlier, for every post P_i, a corresponding set of comments is also defined, symbolized as $\mathcal{C}_i = \{C_{i,j}\}_{j=1}^{M_i}$, where M_i refers to the number of comments tied to the post P_i , we have fixed $M$to be the first 150 comments only for all the posts multiplied by the seven emotions generated by the emotion extraction component model $\boldsymbol{Emo}$. Additionally, the Sentence BERT embedding $\boldsymbol{Emb}_P$ would give us 768 inputs. Those two are the essence of our proposed model named **Code RED**. Furthermore, we randomly split each dataset into 80% training set and a 20% testing set.

The Code RED model integrates emotion-based features and BERT embeddings through a two-part network. Initially, the model uses a three-layer feed-forward network to transform emotion-based features into a 100-dimensional representation. This is then concatenated with 768-dimensional BERT embeddings, forming an 868-dimensional vector. This vector is further transformed via fully connected layers into a 100-dimensional space, which is finally mapped to a binary dimension, indicating class probability through a *CrossEntropyLoss* function. This architecture allows nuanced processing of high-dimensional input data. Our research drew from existing studies to establish baselines for stress classification using **post BERT embedding**, inspired by [14, 19] work, and emotion classification of posts **posts' emotions**, based on the studies by [1, 13, 20]. We also adopted principles [22] **Dual Emotions**, originally used for fake news detection, due to its consideration of comment emotions for classification, highlighting the importance of integrating emotional factors in our social media stress detection techniques.

[8] github.com/Zeyad-o/CODE-RED

4.3 Experimental Results

Analysis of Reactive Emotions Difference (RED) Features Contribution to Prediction To answer (RQ1), we analyze how the Reactive Emotions Difference (RED) feature, i.e. (each individual emotion) correlates to the prediction. We performed a correlation analysis for all of Ekman's basic emotions [5] in relation to the stress label of the post. As shown in Table 2, we considered the emotions derived from our model. However, for this experiment, RED would be based on the absolute value of the difference between the post and the average emotions of the comments.

Our analysis revealed that the reactive emotional difference for some emotions, such as fear and sadness, shows a consistent positive correlation with stress, whereas joy shows a consistent negative correlation. Conversely, disgust and surprise exhibit the weakest correlations. These findings align with established definitions of stress and previous research [1, 20]. However, although anger is conventionally associated with stress, our results indicate that within the context of social media platforms, the discrepancy in the expression of anger between posts and comments does not serve as a strong indicator of stress. Our findings provide a fresh perspective on the relationship between the emotions of stressful posts and their comments. This hints at the possibility of more complex patterns of emotional responses on digital platforms, underscoring the unique dynamics of online emotional interactions.

Features	**Dreadit [19]**	**Rastogi et al.** [14]
Anger	0.081	-0.1308
Disgust	0.0305*	0.1762
Fear	0.193	0.2258
Joy	-0.2212	-0.5622
Sadness	0.1593	0.2016
Surprise	-0.0471	-0.1096

Table 2. The Pearson correlation coefficient for the RED emotional features, the asterisk * has p-value in the range (0.10 - 0.15)

Benchmarking the Effectiveness of RED in Stress Recognition The goal of the experiment is to answer (RQ2) Can we use reactive emotions difference to detect stress, and how does it compare to the baselines?

It is observed as detailed in Table 3, that our novel model, "Code RED", has achieved impressive results, as demonstrated in the Dreadit dataset [19]. It has the highest accuracy of 77.9% among the evaluated models, and an F1 score of 79.2%, which is also the top score. Moreover, when tested on the Rastogi et al. dataset [14], "Code RED" proves its superiority across all performance metrics. It reaches an outstanding accuracy of 92.4%, an F1 score of 93%. We also notice that Rastogi et al. have higher classification performance compared

to Dreadit. We believe that this might be contributed to the dataset curation and annotation process. As mentioned in the dataset section.

Model	**Dreadit** [19]				**Rastogi et al.** [14]			
	Accuracy	F1	Precision	Recall	Accuracy	F1	Precision	Recall
Posts' BERT Emb	76.9	79.1	80.6	77.7	88.8	90	90	**89.8**
Posts' Emotions	73.4	77.6	**85.2**	71.3	80	83	87	79.3
Dual Emotions	72.5	72.4	72.4	72.8	85.8	85.4	85.7	85.2
Code RED	**77.9**	**79.2**	81	**78**	**92.4**	**93**	**97**	89.2

Table 3. Stress Classification Performance for Our Model vs the Baselines.

Our model exhibited high reliability and effectiveness in stress level detection, as evidenced by its superior performance across key metrics, including a balanced precision-recall, reflected in high F1 scores (Table 3). Its capacity to discern emotional discrepancies between posts and corresponding comments highlights its potential utility in applications requiring accurate identification and minimization of false positives.

5 Conclusion and Future Work

Unaddressed stress can have debilitating effects, highlighting the importance of early detection and intervention. Social media posts and interactions have emerged as valuable resources for identifying signs of stress. In our study, we harnessed the distinct reactivity in emotional responses as an auxiliary feature for stress classification. Our analysis showed that some reactive emotional difference features present a consistent correlation to fundamental stress emotions, such as sadness and fear, as well as diminished joy. Nonetheless, other emotions are not as distinctive. We also showcased the novelty and effectiveness of our approach and proven by its superior performance, achieving impressive accuracy and F1 scores. Our study underscores the complexity of emotion-stress correlations in social media posts and comments. They also highlight the need for further research into understanding and leveraging these intricate relationships for mental health applications. For future work, we suggest investigating the usage of Large Language Models (LLM) as a social support channel, comparing its similarity to real-world comments, and fine-tuning LLMs to represent more accurate social support.

References

1. Alghamdi, Z., Kumarage, T., Karami, M., Alatawi, F., Mosallanezhad, A., Liu, H.: Studying the influence of toxicity and emotion features for stress detection on social media. In: European Conference on Social Media. vol. 10, pp. 42–51 (2023)

2. Billings, A.G., Moos, R.H.: Coping, stress, and social resources among adults with unipolar depression. Journal of personality and social psychology **46**(4), 877 (1984)
3. Breslau, N., Schultz, L., Peterson, E.: Sex differences in depression: a role for preexisting anxiety. Psychiatry research **58**(1), 1–12 (1995)
4. Cohen, S., Hoberman, H.M.: Positive events and social supports as buffers of life change stress 1. Journal of applied social psychology **13**(2), 99–125 (1983)
5. Ekman, P.: Are there basic emotions? Psychological Review, 99(5):550–553. (1992)
6. Faravelli, C., Pallanti, S.: Recent life events and panic disorder. The American journal of psychiatry (1989)
7. Goeders, N.E.: The impact of stress on addiction. European Neuropsychopharmacology **13**(6), 435–441 (2003)
8. Khan, A., Ali, R.: Stress detection from twitter posts using lda. International Journal of High Performance Computing and Networking **16**(2-3), 137–147 (2020)
9. Leavy, R.L.: Social support and psychological disorder: A review. Journal of community psychology **11**(1), 3–21 (1983)
10. Lin, H., Jia, J., Guo, Q., Xue, Y., Li, Q., Huang, J., Cai, L., Feng, L.: User-level psychological stress detection from social media using deep neural network. In: Proceedings of the 22nd ACM international conference on Multimedia (2014)
11. Lin, H., Jia, J., Qiu, J., Zhang, Y., Shen, G., Xie, L., Tang, J., Feng, L., Chua, T.S.: Detecting stress based on social interactions in social networks. IEEE Transactions on Knowledge and Data Engineering **29**(9), 1820–1833 (2017)
12. Matthews, K.A., Katholi, C.R., McCreath, H., Whooley, M.A., Williams, D.R., Zhu, S., Markovitz, J.H.: Blood pressure reactivity to psychological stress predicts hypertension in the cardia study. Circulation **110**(1), 74–78 (2004)
13. Nijhawan, T., Attigeri, G., Ananthakrishna, T.: Stress detection using natural language processing and machine learning over social interactions. Journal of Big Data **9**(1), 1–24 (2022)
14. Rastogi, A., Liu, Q., Cambria, E.: Stress detection from social media articles: New dataset benchmark and analytical study. In: (IJCNN). pp. 1–8. IEEE (2022)
15. Sher, L.: The impact of the covid-19 pandemic on suicide rates. QJM: An International Journal of Medicine **113**(10), 707–712 (2020)
16. Slopen, N., Kontos, E.Z., Ryff, C.D., Ayanian, J.Z., Albert, M.A., Williams, D.R.: Psychosocial stress and cigarette smoking persistence, cessation, and relapse over 9–10 years: a prospective study of middle-aged adults in the united states. Cancer Causes & Control **24**, 1849–1863 (2013)
17. Sowles, S.J., McLeary, M., Optican, A., Cahn, E., Krauss, M.J., Fitzsimmons-Craft, E.E., Wilfley, D.E., Cavazos-Rehg, P.A.: A content analysis of an online pro-eating disorder community on reddit. Body image **24**, 137–144 (2018)
18. Thelwall, M.: Tensistrength: Stress and relaxation magnitude detection for social media texts. Information Processing & Management **53**(1), 106–121 (2017)
19. Turcan, E., McKeown, K.: Dreaddit: A reddit dataset for stress analysis in social media. EMNLP-IJCNLP 2019 p. 97 (2019)
20. Turcan, E., Muresan, S., McKeown, K.: Emotion-infused models for explainable psychological stress detection. In: Proceedings of NAACL (2021)
21. Yudkin, J.S., Kumari, M., Humphries, S.E., Mohamed-Ali, V.: Inflammation, obesity, stress and coronary heart disease: is interleukin-6 the link? Atherosclerosis **148**(2), 209–214 (2000)
22. Zhang, X., Cao, J., Li, X., Sheng, Q., Zhong, L., Shu, K.: Mining dual emotion for fake news detection. In: Proceedings of the web conference 2021 (2021)