

StressSGCL: A Stress-Specific Embeddings Learning Approach using Contrastive Learning over Skip-Gram Model

Muhammad Abulaish¹, Rumjot Kaur¹, Amit Kumar Sah¹

¹Department of Computer Science and Engineering, South Asian University, New Delhi
abulaish@sau.ac.in, rumjotkaur219@gmail.com, amitcsrs@students.sau.ac.in

Abstract

In this paper, we present *StressSGCL*, a novel approach that produces stress-specific word embeddings to enhance stress detection in social media posts. The proposed approach improves contextual representation by integrating the skip-gram model with transformers-based Pre-trained Language Models (PLMs). Subsequently, it refines these embeddings through supervised contrastive learning, which focuses on the differentiation of stress-related emotional content. The *StressSGCL* is evaluated using two PLMs (BERT and MentalBERT) on three benchmark datasets from Twitter and Reddit. Empirical evaluation employing the 10-fold cross-validation demonstrates that the stress-specific embeddings produced by *StressSGCL* consistently outperform various ablation baselines and existing stress detection methods. The *StressSGCL* advances mental health text analysis by demonstrating how static embedding techniques can effectively supplement dynamic PLMs to improve performance when dealing with domain-specific problems such as stress detection.

1 Introduction

Stress is the body’s reaction to demanding or challenging situations, varying subjectively from person to person. These events trigger hormonal changes, leading to diverse psychological and physiological alterations in our body—the body’s ability to adapt to stress aids in coping with short-term challenges. Stress can be categorized as physical, psychological, psychosocial, and psycho-spiritual¹. Modern life’s rapid pace, finances, work, and society are familiar sources of increased daily stress. Today, people are more concerned about their careers than health and are hardly interested in visiting a psychiatrist. The COVID-19 pandemic has affected the mental health of the population like anything. According to the American Psychological Association’s 2020 stress report², 67% of Americans agreed to increased stress during the pandemic while the young generation, or Gen Z, was reported to be more stressed as they faced the uncertain future, on average, 6.1 out of 10. In India, one in every third of Indians is struggling with stress

and anxiety³. The advent of online social networks has provided individuals with platforms to express their feelings and thoughts, including their experiences with stress. Social networks help them meet people undergoing similar conditions; for instance, the Reddit⁴ social media platform hosts communities dedicated to mental health discussions- r/MentalHealth, r/Depression, r/Anxiety, r/Suicide, and many more. This shift towards online expression offers researchers a rich source of time-series data for investigating stress and mental health disorders. Prolonged exposure to stress leads to chronic stress, causing various issues like mental health problems, including depression and anxiety, a weakened immune system, and potentially fatal outcomes if left unmitigated. Hence, detecting symptoms and signs of stress before it becomes a severe problem is important.

Researchers have applied different Machine Learning (ML) and Deep Learning (DL) approaches to detect stress from online social networks. Studies done by Wang et al. (2023); Ilias, Mouzakitis, and Askounis (2023) highlight that the stress detection field’s main focus has been utilizing word embeddings, including Word2Vec and BERT, to represent each user post. These embeddings can learn the syntactic and semantic relationships among the words of a sentence. On the contrary, for feature selection, one has to identify the key factors that play a significant role in stress detection, as mentioned earlier in studies done by Lin et al. (2014b,a). Recently, transformers-based language models, including BERT, RoBERTa, and GPT, have revolutionized the Natural Language Processing (NLP) field, providing dynamic embeddings that are more contextually and semantically enriched, unlike the traditional language models generating static embeddings, including Skip-Gram, Continuous-Bag-of-Words, Global Vectors for Word Representation (GloVe) and FastText. Wang, Cui, and Zhang (2021) studies that the transformers-based language models fail to capture context more effectively despite their popularity.

In contrast, static embeddings have demonstrated commendable performance across various downstream NLP

Copyright © 2025, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

¹<https://www.mentalhealth.com/library/types-of-stressors>

²<https://www.apa.org/news/press/releases/stress/2020/sia-mental-health-crisis.pdf>

³<https://frontline.thehindu.com/news/explained-why-younger-generations-in-india-seem-so-stressed-climate-change-war-conflict-social-media-mental-health/article67792037.ece>

⁴<https://www.reddit.com/>

tasks. For instance, in mental health research, researchers have explored diverse approaches, including utilizing BiLSTM with Word2Vec or FastText embeddings, LSTM with GloVe embeddings, or trainable embedding layers. Researchers have undertaken significant efforts to bridge the gap between static and dynamic embeddings. In this direction, Wang, Cui, and Zhang (2021) proposed to generate static embeddings from BERT. Notably, Vashishth et al. (2019) attempted to enhance Skip-Gram embeddings with Graph Convolutional Networks (GCN) to integrate sentence semantics into the output embeddings. These models have demonstrated improved accuracy across various NLP tasks like Parts-of-Speech tagging, Named Entity Recognition (NER), and different word similarity tasks. This motivates us to apply the combination of dynamic and static embeddings in the field of stress detection.

Contrastive learning is a self-supervised technique that creates representations by comparing data points with the primary aim of grouping similar samples closer together in a shared embedding space while ensuring that dissimilar samples are placed further apart. Contrastive learning is a massive success in the field of NLP (Gao, Yao, and Chen 2021) and computer vision (Chen et al. 2020). Its core objective is to produce distinct and effective representations for different categories within a dataset. Khosla et al. (2020) proposed Supervised Contrastive Learning (SCL) which makes better use of label information, resulting in improved performance on supervised tasks. According to psychological theory, emotion, and stress move parallel (Du et al. 2018). Stress can lead to negative emotions such as depression, anxiety, anger, fear, and frustration. Detecting stress through emotions is common because the higher the stress level, the more intense the negative emotions experienced. We used this strong connection to identify stressed posts and incorporated it into our model through contrastive learning to learn stress-specific embedding.

In this paper, we propose *StressSGCL*, a stress detection approach that learns stress-specific embeddings employing SCL over the Skip-Gram model from social media posts. To the best of our knowledge, the *StressSGCL* is a novel work in stress detection, explicitly focusing on English language social media posts learning stress-specific embeddings employing Skip-Gram and SCL. *StressSGCL* contributes to incorporating additional knowledge into the state-of-the-art deep learning models, enabling them to detect stress more accurately. Besides, we evaluated the *StressSGCL* on datasets from various social media platforms, namely Twitter (X) and Reddit, ensuring the model's generalizability.

2 Related works

In recent years, the diagnosis of mental disorders has gained significant importance, leading researchers to delve into the analysis of social media data to enhance predictive accuracy. For instance, De Choudhury and De (2014) explored the relationship between mental health issues and social media expressions, while Yates, Cohan, and Goharian (2017) examined online forums to identify signs of depression and assess the risk of self-harm. Notably, Lokala et al. (2022)

classified mental health symptoms specifically for heart patients, taking gender differences into account. In addition, Wang et al. (2020) utilized visual-textual multimodal learning techniques to detect symptoms of depression effectively.

Moreover, the investigation of suicide risk through social media platforms has been extensive, employing various methodologies (De Choudhury et al. 2016; Jashinsky et al. 2014; Poulin et al. 2014; Coppersmith et al. 2018). Recognizing the complexity of stress detection, researchers have also considered a wide range of factors in their diagnostic approaches, which include both physiological data (Raij et al. 2010; Lu et al. 2012) and social media interactions (Wang et al. 2023; Thelwall 2017).

Research on stress can generally be categorized into two main areas: tweet-level stress detection and user-level stress detection.

2.1 User-level stress detection

At user-level stress detection, researchers typically analyze the time series of user posts to discern patterns indicative of stress. A high-level representation is learned for each social media user, which is then fed into a classification model. Lin et al. (2014b) focused on detecting Twitter (X) users' user-level psychological stress state by leveraging multimodal user-defined features at both the tweet and user levels. They employed a cross-media auto-encoder to merge cross-media data and utilized a convolutional neural network to combine tweet and user-level features. Lin et al. (2017) integrated the textual, visual, and social attributes with the social interaction features of Twitter (X) users to enhance stress detection. Wang et al. (2023) emphasized the significance of stressors and stressful emotions in stress detection. They focused on extracting posts from the Sina Weibo social media platform containing stressful emotions and stressors. Employing a contrastive objective, they fine-tuned the BERT embeddings to generate stress-specific representations, achieving notable advancements in stress detection capabilities.

2.2 Tweet-level stress detection

At tweet-level stress detection, researchers analyze the information within individual user posts to identify stress indicators. Lin et al. (2014a) adopted a comprehensive approach by utilizing user-defined multimodal attributes, including linguistic, visual, and social aspects, to categorize users into one of five stress categories. Ilias, Mouzakitis, and Askounis (2023) incorporated linguistic features of posts into transformer-based models to detect stress and depression. Lin et al. (2016) extracted user-defined features based on stressor events and subjects and word-level vector representation, then fed them to a hybrid multi-task model to detect stressor events and subjects. Turcan, Muresan, and McKeown (2021) proposed an emotion-infused BERT model to detect stress. Turcan and McKeown (2019), authors explored various machine learning and deep learning classifiers to detect stress using linguistic features and word embeddings. Thelwall (2017) constructed an extensive lexicon to quantify users' stress levels and relaxation states based on linguistic cues, contributing to the development of stress detection models.

Researchers have predominantly detected stress using user-defined features or word embeddings (mainly transformers-based language models), with some studies exploring combinations of both approaches. Despite the efficacy of transformers-based language models, there have been efforts to enhance their effectiveness, considering their limitations. Bommasani, Davis, and Cardie (2020) took the contextualized representation from BERT and built a static embedding by averaging certain word representations. In another approach, Wang, Cui, and Zhang (2021) infused BERT embeddings into a Skip-Gram model to capitalize on both techniques’ strengths, potentially enhancing the accuracy and robustness of static embeddings. In this work, we focus on short-term stress where the focus is on a single post of a user, and the state of a post defines the state of that particular user. However, the proposed approach can also be extended to chronic stress detection, which requires analyzing user posts’ time series.

3 Proposed approach

This section introduces the proposed approach, *StressSGCL*, designed for detecting stress in social media texts. It utilizes a Skip-Gram model based on the SCL framework in conjunction with transformers-based Pre-trained Language Models (PLMs). Our goal is to develop stress-specific embeddings by first integrating the Skip-Gram model with the embeddings generated by the transformers-based PLM. This integration improves the estimation of context words within a specified window, resulting in target words that are both semantically and syntactically richer. We then incorporate stress-specific features into these embeddings through the SCL process. The complete architecture of the proposed model is illustrated in Figure 1.

As a result, this approach enables the model to better understand stress-related emotions by employing an SCL task. Section 3.1 explains how we extract dynamic embeddings from transformers-based PLMs. After that, Section 3.2 explains the learning of better contextualized static embeddings by integrating dynamic embeddings with Skip-Gram. Then, Section 3.3 explains how we apply SCL over the embeddings learned in Section 3.2 to learn stress-specific embeddings capturing the relation of words with stress or non-stress category. Finally, Section 3.4 presents the classifier we use to validate the effectiveness of the stress-specific embeddings learned.

3.1 Embeddings extraction from transformers-based PLM

We validate *StressSGCL* using two transformers-based PLMs, namely, (i) BERT – Bidirectional Encoder Representations from Transformers (Devlin et al. 2019), and (ii) MentalBERT – Publicly Available Pretrained Language Models for Mental Healthcare (Ji et al. 2022). BERT is a transformers-based language model trained on a general (non-specialized) corpus, namely BooksCorpus and English Wikipedia. At the same time, MentalBERT is based on BERT and pre-trained on the corpus from Reddit, which

constitutes discussions focused on the mental health domain. MentalBERT is based on BERT and follows the same text preprocessing as BERT. We used BERT base uncased and MentalBERT base uncased models.

Given a sentence S with n words, $S = \{w_1; w_2; w_3; \dots; w_n\}$, we retrieve dynamic word embeddings based on the context using BERT for each word. Towards this, firstly, we add tokens $[CLS]$ and $[SEP]$ at the beginning and the end of S , respectively, as given by Equation 1.

$$S = [CLS]; w_1; w_2; w_3; \dots; w_n; [SEP] \quad (1)$$

We further tokenize S using WordPiece tokenizer⁵ and transform it to a vector, a combination of positional, segment and token embedding. Hence, each word w_i in S is represented as w_i , sum of token (TE_{w_i}), segment (SE_{w_i}) and positional embeddings (PE_{w_i}), given by Equation 2.

$$w_i = TE_{w_i} + PE_{w_i} + SE_{w_i} \quad (2)$$

Further, $w_1; w_2; w_3; \dots; w_n$ are fed to the transformers-based PLM model (*PLM*) to generate the output embeddings, u_i , such that $u_i \in \mathbb{R}^d$, where, $d = 768$.

$$u_{w_i} = PLM(w_i) \quad (3)$$

When dealing with subwords generated by the WordPiece tokenizer, we use a mean pooling operation to obtain the final embedding for a word.

3.2 Skip-Gram embeddings

Skip-Gram model (Mikolov et al. 2013) aims to learn the co-occurrence between a target word and the words in the context window to generate the word embeddings for each word. To incorporate the embeddings from transformers-based PLM into the Skip-Gram model in a sentence, we replace target word embeddings with its respective embeddings learned using the PLM, which we refer to as SG_{PLM} embeddings.

Given a sentence $S = \{w_1; w_2; w_3; \dots; w_n\}$, let w_i be the target word, c be window size, then $\{w_{i-c}; \dots; w_{i-1}; w_{i+1}; \dots; w_{i+c}\}$ are the context words. Mathematically, the objective of the Skip-Gram model is to maximize the probability of context words for a given target word w_i , illustrated in Equation 4.

$$\frac{1}{n} \times \prod_{i=1}^n \prod_{c: j \neq i} \log p(w_{i+j} | w_i) \quad (4)$$

We substitute the target word w_i with its corresponding transformers-based PLM embeddings u_i . The primary purpose of replacing the target word embeddings w_i in the skip-gram model with contextualized embeddings u_i generated by the PLM is to harness the rich contextual information captured by u_i . Unlike static embeddings, which treat words in isolation, transformers-based PLM-derived embeddings

⁵https://keras.io/api/keras_nlp/tokenizers/word_piece_tokenizer/

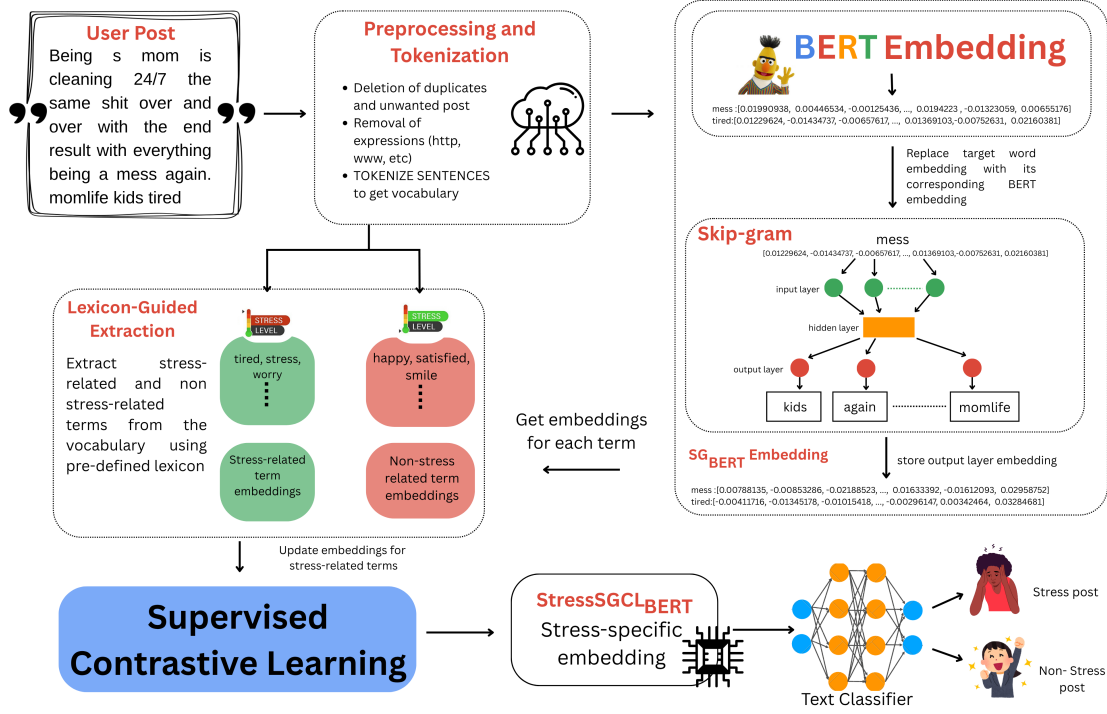


Figure 1: Architecture of the proposed *StressSGCL* framework integrating BERT (a transformers-based PLM) with Skip-Gram and SCL to generate stress-specific embeddings. BERT shown here can be substituted with MentalBERT.

are dynamic and encapsulate the semantic and syntactic context of a word within its sentence. This enables the skip-gram model to utilize more nuanced, context-aware representations of words, potentially enhancing its ability to accurately predict context words. Let V be a matrix that denotes the embeddings of the output layer of the Skip-Gram model and represents the final output static embeddings.

$$V = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \quad (5)$$

In Equation 5, $v_i \in \mathbb{R}^d$; $d = 768$ and n is vocabulary size.

To learn the value of v_i for a context word, the revised goal of Skip-Gram, as determined in Equation 6, is to reduce the noise contrastive loss function through negative sampling where k defines the number of noise words considered for negative sampling.

$$L_{SG} = \sum_{j=1}^c \sum_{i=1}^c \log(v_i^T; u_j) + \sum_{m=1}^k \sum_{i=1}^c E_{w_m \sim P(w)} \log(v_m^T; u_j) \quad (6)$$

In the proposed approach, c as well as k is set to 5. $P(w)$ defines the uniform noise distribution from which noise words are sampled.

In Equation 7, the function *Skip Gram* embodies the Skip-Gram model, where both the context and target word are provided as inputs, yielding the resultant learned embeddings as its output.

$$v_i^o = \text{Skip Gram}(v_i; u_i) \quad (7)$$

where, v_i^o is the final output embeddings learnt against every v_i . Equation 8 represents the learned embeddings matrix V^o .

$$V^o = \begin{bmatrix} v_1^o \\ v_2^o \\ \vdots \\ v_n^o \end{bmatrix} \quad (8)$$

3.3 Supervised Contrastive Learning (SCL)

Contrastive learning is an approach to learning embeddings based on semantic similarity. The objective of contrastive learning is to learn representation embeddings such that similar samples (positive pairs) come close to each other. In contrast, dissimilar samples (negative pairs) are pushed far apart in the embedding space. SCL is a type of contrastive learning that leverages ground truth labels to refine the selection of positive and negative samples and allow for multiple positives in training. The overall objective can be given by Equation 9. Here, if C refers to a class of the dataset and x_a is an anchor point for which we need to learn an embedding, such that ($x_a \in C$); and there is an instance x_p of the dataset,

where $(x_p \in C \text{ and } x_p \notin x_a)$ forms a positive pair with x_a while an instance x_n of the dataset, where $(x_n \notin C)$ forms a negative pair with x_a .

$$\text{sim}(x_a; x_p) \gg \text{sim}(x_a; x_n) \quad (9)$$

where, $\text{sim}()$ is a similarity function.

Our proposed approach frames the Stress-related Emotion Distinction (SED) as a contrastive task. We utilize SCL to integrate stress-specific features into the SG_{PLM} embedding. To achieve the objectives of SCL, we employ the InfoNCE loss function (Oord, Li, and Vinyals 2018). This SED task enhances the SG_{PLM} embedding, making it more attuned to stress-related emotions. We identify these emotions using the TensiStrength lexicon. We discuss the process of identifying stressful lexicons and stressful lexicon-based contrastive learning as follows.

- (a) **Stressful emotion identification:** In psychology, the stressor is the cause of stress (Wang et al. 2023), and stress-related emotion is generally the negative emotions that arise from stress (Du et al. 2018). For example, ‘‘I am stressed because of my exams’’; here, ‘‘exam’’ is a stressor, and ‘‘stressed’’ is a stress-related emotion. The term ‘‘exam’’ will not be a stressor without the presence of stress-related emotions in the document. Considering this, our first task is to identify the presence of both the stressor and stress-related emotions in the documents.

Given that our datasets comprise documents collected based on hashtags or subreddits, it’s imperative to ensure that each document contains a stress-related emotion. To achieve this, we employed the TensiStrength lexicon proposed by Thelwall (2017), which is a system to measure the intensity of stress and relaxation conveyed in social media posts, particularly in the transportation context. This method utilizes a lexicon-based approach and a set of rules to detect direct and indirect expressions of stress or relaxation. Each term within the lexicon is assigned a magnitude ranging from 1 (very weak) to 5 (very strong), denoting its power in conveying stress or relaxation.

To employ the TensiStrength lexicon effectively, we exclude the transportation and relaxation terms, resulting in a lexicon size of 2,064 terms. This ensured a focused analysis of stress-related emotions within our datasets. Subsequently, based on the cosine similarity score, we validate if the document contains stress-related emotions akin to those in the lexicon. We set the similarity threshold of 0.80 to guarantee a close match between the stress-related emotions expressed in the document and those recorded in the lexicon.

- (b) **Stressful lexicon-based contrastive learning:** Let \mathcal{L} represent the set of words in the TensiStrength lexicon, set \mathcal{A} represent the antonyms of words in set \mathcal{L} , and set V_S and V_{ns} represent the vocabulary of documents labeled as stress and non-stress respectively. We define two sets – (i) P with stress-related emotions given by Equation 10, and (ii) N with non-stress-related emotions given by Equation 11. Here, $P \setminus N = ?$.

$$P = \{v_{w_i} \mid w_i \in V_S \text{ and } w_i \in \mathcal{L}\} \quad (10)$$

$$N = \{v_{w_i} \mid w_i \in V_{ns} \text{ and } w_i \in \mathcal{L}\} \quad (11)$$

Based on our objective, we only update the embedding of set P with SCL. Thus, each word in set P becomes an anchor point. The objective of the SED task is computed as per the Equation 12 where v_i^o denotes the stress-related emotion, v_{pi}^o denotes the positive sample from set P , v_{nj}^o denotes negative sample from set N , r is cardinality of set N and β is the temperature hyperparameter.

$$\text{SED} = \log \frac{\exp(v_i^{oT} v_{pi}^o)}{\sum_{j=0}^r \exp(v_i^{oT} v_{nj}^o)} \quad (12)$$

To implement SCL, we employ a two-layer neural network with 768 neurons without any activation function. We retrieve enhanced stress-specific embeddings based on StressSGCL, which we further employ for stress detection, discussed in the following section.

3.4 Classification model

StressSGCL performs stress detection employing the TextCNN classifier (Chen 2015) with max pooling to evaluate the efficacy of stress-specific embeddings learned. To this, TextCNN utilizes multiple filters to extract various features from the input data. We input a 3-D matrix $\mathbb{R}^{n \times d \times 1}$ corresponding to each document in the dataset where n is the total number of words in the document selected for classification task which in our case is dataset dependent since the TextCNN model accepts fixed size input and $d = 768$ is the dimension of the stress-specific embeddings learned.

4 Experimental setup and results

In this section, we evaluate the effectiveness of our proposed StressSGCL approach for stress detection in social media text. We begin in Section 4.1 by introducing the three benchmark datasets collected from Twitter (X) and Reddit platforms. Section 4.2 then describes the baseline methods and ablation variants used to validate our model’s components. In Section 4.3, we outline the experimental configurations, hyperparameters, and evaluation metrics employed across all experiments. Section 4.4 presents a thorough comparative analysis using both quantitative metrics and qualitative visualizations to demonstrate the superiority of our stress-specific embeddings. All experiments were implemented in Python and executed on a system with a 2.10 GHz Intel(R) Silver(R) processor and 192 GB RAM.

4.1 Datasets

We evaluate StressSGCL on three publicly available benchmark stress detection datasets proposed by Rastogi, Liu, and Cambria (2022) namely, (a) **Twitter Full** (DS₁), (b) **Reddit Title** (DS₂), and (c) **Reddit Combi** (DS₃). The Twitter full dataset was crawled from Twitter (X) using stress-related hashtags like #Stressed, #Fatigue, and #Tired, as well as positive emotion hashtags like #Happy, #Joy, and

Table 1: Statistics of the datasets

Dataset	#Stress	#Non-stress
Twitter Full (DS ₁)	4,534	4,366
Reddit Title (DS ₂)	2,745	2,811
Reddit Combi (DS ₃)	2,745	859

#Blessed. It contains 4,534 stress tweets and 4,366 non-stress tweets, making it a balanced dataset. The Reddit Title dataset includes titles of posts from stress-related and non-stress-related subreddits. It is a balanced dataset with 2,745 stress posts and 2,811 non-stress posts, similar to the Twitter Full dataset in terms of post length. The Reddit Combi dataset includes body text and post titles from stress and non-stress subreddits. It is imbalanced with 2,745 stress posts and 859 non-stress posts, as most bodies contain non-text data like images and videos. Table 1 summarizes the statistics of the three benchmark datasets undertaken in this study.

4.2 Comparison approaches

This section discusses various comparison approaches undertaken in this study to validate the effectiveness of the proposed stress-detection approach.

We compared `StressSGCL` with the following two stress detection approaches.

- (i) **Contrastive Learning of Stress-specific Word Embeddings for Social Media based Stress Detection (SSE)**: Wang et al. (2023) proposed SSE, a BERT-based stress-specific embeddings. For this, they defined three self-supervised contrastive tasks: 1) stressor discrimination task, 2) stressor cluster discrimination task, and 3) stressful emotion discrimination task. With the stress-specific embeddings, they analyzed the user’s time-series posts to detect their stress state. We compared `StressSGCL` with both BERT and MentalBERT versions of SSE defined as `SSEBERT` and `SSEMentalBERT`.
- (ii) **Stress and relaxation magnitude detection for social media texts (TensiStrength)**: Thelwall (2017) proposed TensiStrength, a lexicon-based approach to identify the magnitude of stress and relaxation state for each post, considering the intensity of individual words in the lexicon. To convert it to a classification problem, we considered the maximum value among two magnitudes to identify its state. For instance, if a post has three as stress magnitude and five as relax magnitude, we labeled it as non-stress.

Additionally, we created four ablation-like baselines to validate the effectiveness of the modules undertaken to develop `StressSGCL`. The first two baselines are obtained by removing SCL from `StressSGCL`, namely (i) `SGBERT` and (ii) `SGMentalBERT` while the other two are fine-tuned transformers-based PLMs, namely, (iii) `BERT` and (iv) `MentalBERT`, to assess the individual effectiveness of the two PLMs undertaken to validate the effectiveness of `StressSGCL` as stress-detection methods. We disassociated SCL from `StressSGCLBERT` to come

up with `SGBERT` with an aim to study the improvement that SCL brings in the final output embeddings extracted from the `StressSGCLBERT` model. Similarly, we disassociated SCL from the `StressSGCLMentalBERT` model to develop `SGMentalBERT` to study the improvement that SCL brings to the final output embeddings extracted from the `StressSGCLMentalBERT` model. Towards BERT and MentalBERT, we fine-tuned the pre-trained BERT base model on the datasets such that the dense vector learned corresponding to `[CLS]` tokens corresponding to each document is passed through softmax layer for the classification task.

4.3 Training details

In this section, we discuss the different training hyperparameters undertaken in this study. To learn embeddings using `SGPLM`, we used Adam optimizer with a learning rate of 0.001 and batch size of 64. We implement the infoNCE loss function in the SED task with a temperature hyperparameter set to 0.07. With Adam optimizer, epoch set to 20, and learning rate set to 0.0001, we trained the stress-specific embeddings with the SCL model described in Section 3.3. We evaluated the proposed approach and the comparison approaches using 10-fold cross-validation. Additionally, we split 20% of the training set for validation and report the average performance over the 10-fold. We trained the proposed stress-detection model using the Adam optimizer with a learning rate of 0.001 and binary cross-entropy as a loss function for 100 epochs and a batch size of 64. For TextCNN, We restricted the length of each document to the average length of documents in the dataset, which was 37 for Twitter Full, 17 for Reddit Title, and 80 for Reddit Combi. We evaluate the proposed `StressSGCL` and the comparison approaches, including ablation-like baselines, using four standard evaluation metrics, namely Accuracy, F_1 score of the stress class (F_1), Macro F_1 score ($MacroF_1$), and Receiver Operating Characteristic Curve (ROC Curve)-based analysis. These evaluation metrics are discussed as follows.

- (a) **Accuracy**: It measures the total number of correct predictions made by the classifier, given by Equation 13.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (13)$$

Here, TP, TN, FP, and FN are true positives, true negatives, false positives, and false negatives, respectively.

- (b) **F_1 -Score**: It measures the harmonic mean value of precision and recall, given by Equation 14.

$$F_1 = 2 \frac{Precision \cdot Recall}{Precision + Recall} \quad (14)$$

Here, $Precision = \frac{TP}{TP+FP}$, and $Recall = \frac{TP}{TP+FN}$. We calculate the F_1 score value of the stress class.

- (c) **Macro F_1 -Score**: It computes class averaged F_1 -Score value and is given by Equation 15, where $F_{1\ stress}$ gives the F_1 score of stress class and $F_{1\ non\ stress}$ gives the F_1 score of the non-stress class.

$$MacroF_1 = \frac{F_{1\ stress} + F_{1\ non\ stress}}{2} \quad (15)$$

- (d) **ROC Curve:** A ROC curve, or receiver operating characteristic curve, serves as a graphical representation of a classification model’s performance across all possible thresholds. This visualization highlights two key metrics: (i) the True Positive Rate (TPR), also known as recall, and (ii) the False Positive Rate (FPR), which is calculated using the formula $FPR = \frac{FP}{TN+FP}$. By plotting these two parameters against one another, the ROC curve effectively illustrates the model’s performance and provides an Area Under the ROC Curve (AUC), offering a comprehensive evaluation of the model’s effectiveness across various classification thresholds.

4.4 Comparative performance analysis

In this section, we present the comparative study of the proposed models, namely, $StressSGCL_{BERT}$ and $StressSGCL_{MentalBERT}$ against comparison approaches discussed in Section 4.2 over the three benchmark datasets discussed in Section 4.1. Additionally, we report the comparative results obtained in Table 2. We mention $StressSGCL_{PLM}$, SG_{PLM} , and SSE_{PLM} when referring to the case of both BERT and MentalBERT.

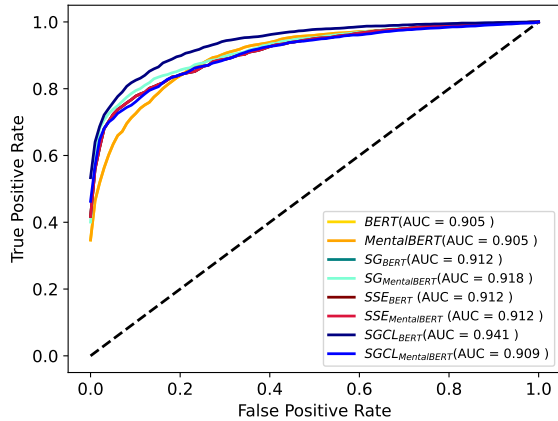
- (a) **Twitter Full:** We observe from Table 2 that integration of SCL with the baseline SG_{BERT} notably enhanced performance as $StressSGCL_{BERT}$ achieved remarkable performance, with the highest F_1 score of 91.03%. This was followed by $StressSGCL_{MentalBERT}$, SSE_{PLM} , and $SG_{MentalBERT}$ all of which achieved F_1 score above 90%. Notably, the SED did not demonstrate the same effectiveness in enhancing the embeddings generated by $SG_{MentalBERT}$. Despite this, it exhibited comparable performance against the SSE_{PLM} stress detection method. Notably, the SG_{PLM} model outperformed the fine-tuned MentalBERT and BERT. This underscores the efficacy of integrating transformers-based models with static embeddings like Skip-Gram. Moreover, TensiStrength performed poorly. Fine-tuned BERT attained the lowest performance with a F_1 score of 77.31%, while fine-tuned MentalBERT demonstrated superior performance compared to fine-tuned BERT. This improvement can be attributed to MentalBERT being initially pre-trained on mental health-related datasets, highlighting the importance of domain-specific pre-training in achieving optimal performance in mental health-related tasks. Overall, $StressSGCL_{PLM}$ outperformed the rest of the comparison approaches on the Twitter Full dataset in terms of all Accuracy, F_1 , and $MacroF_1$.
- (b) **Reddit Title:** From Table 2, we observed that incorporating SCL in SG_{PLM} significantly improved its performance across the Reddit Title dataset. Notably, $StressSGCL_{PLM}$ achieved an impressive F_1 score of approximately 93%, preceded by SSE_{BERT} . $SSE_{MentalBERT}$ demonstrated performance akin to SG_{PLM} , underscoring the effectiveness of SG_{PLM} in this context. Comparative analysis revealed that SG_{PLM} outperformed fine-tuned versions of MentalBERT and BERT by margins of approximately 2% and 9%, respectively, in terms of F_1 score. Like in the case of the Twitter

Full dataset, fine-tuned MentalBERT showcased superior performance compared to fine-tuned BERT. Overall, TensiStrength demonstrated under-performance on this dataset. One possible reason could be that Reddit Titles often represent condensed versions of the original posts, and the effectiveness of TensiStrength depends upon the nature of documents, with documents rich in stress-related terms being problematic. Instances such as “School work”, “Can’t anymore” and similar documents in Reddit Title pose challenges for TensiStrength in accurately detecting the correct class. Overall, $StressSGCL_{PLM}$ outperformed the rest of the comparison approaches on the Reddit Title dataset in terms of all Accuracy, F_1 , and $MacroF_1$.

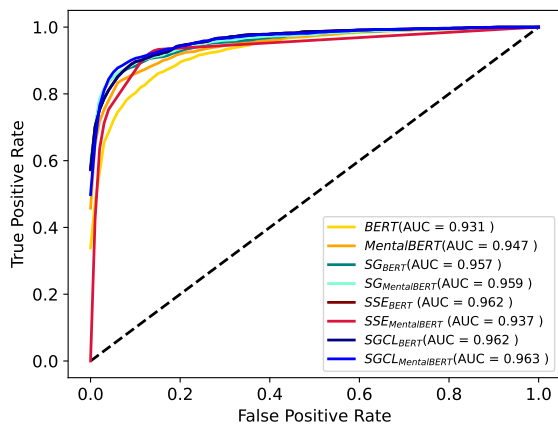
- (c) **Reddit Combi:** The $MacroF_1$ score is a robust comparison metric in evaluating model performance on the imbalanced Reddit Combi dataset. Table 2 shows that $StressSGCL_{MentalBERT}$ performed the best of all in terms of $MacroF_1$ score, achieving an impressive value of 86.01%. which was followed by $StressSGCL_{BERT}$ and SG_{BERT} exhibiting strong performance with $MacroF_1$ score of approximately 85%, while SSE_{PLM} and $SG_{MentalBERT}$ achieved scores of 84%. Notably, both SG_{BERT} and $SG_{MentalBERT}$ surpassed their base counterparts, fine-tuned BERT and fine-tuned MentalBERT, by significant margins, achieving improvements of 14% and 8%, respectively, in terms of $MacroF_1$. Once again, fine-tuned MentalBERT demonstrated superior performance compared to fine-tuned BERT. Conversely, the lowest $MacroF_1$ was observed with TensiStrength. Overall, $StressSGCL_{PLM}$ outperformed the rest of the comparison approaches on the Reddit Combi dataset except in terms of F_1 score where SG_{BERT} performed the best.

ROC Curve-based comparative performance analysis

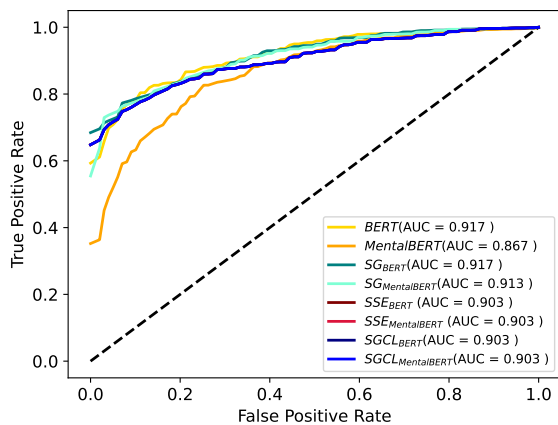
Figure 2 presents the average Area Under the receiver operating characteristic Curve (AUC) value of the proposed $StressSGCL$ and the comparison approaches over the three benchmark datasets from 10-fold cross-validation. The proximity of these curves to the top-left corner of the plot signifies the robustness and effectiveness of the models under evaluation. The Reddit Title dataset showcased the highest performance among the three datasets, followed by the Twitter Full and Reddit Combi datasets with the highest average AUC value, 0.963. $StressSGCL$ occupies the leftmost position on the ROC curves in Figure 2a and Figure 2b, indicating its superior performance across both Twitter and Reddit Title datasets. $StressSGCL_{BERT}$ achieved an average AUC of 0.941 on the Twitter Full dataset, while on the Reddit Title dataset $StressSGCL_{MentalBERT}$ attained an average AUC of 0.963. In contrast, Figure 2c illustrates comparable performance among most models, except the fine-tuned MentalBERT model. Over Reddit Combi dataset, $StressSGCL_{BERT}$, $StressSGCL_{MentalBERT}$, SSE_{BERT} and $SSE_{MentalBERT}$ achieved an average AUC value of 0.903.



(a) Twitter Full



(b) Reddit Title



(c) Reddit Combi

Figure 2: ROC curve-based performance comparison of StressSGCL across three datasets.

Visualization and analysis of embedding spaces To evaluate the effectiveness of our StressSGCL_{PLM} in generating stress-specific embeddings, we employed t-distributed Stochastic Neighbor Embedding (t-SNE) for dimensionality reduction. This technique projects our high-dimensional embeddings into a 2D space while preserving relative distances between tokens, allowing us to visualize semantic relationships that support our primary claim that contrastive learning produces more discriminative stress-aware representations. As demonstrated in the appendix in Figure 3(c), Figure 3(d), Figure 4(c), Figure 4(d), Figure 5(c) and Figure 5(d), StressSGCL_{PLM} generates a distinctive embedding space with clear separation between two major clusters: one predominantly consisting of neutral or emotionally positive tokens, and another comprising stress-related terms such as “anxiety,” “worry,” “overwhelmed,” and “exhausted.” This separation directly validates our core contribution—the creation of embeddings that specifically distinguish stress-related language patterns from general vocabulary. The visualization results align with our quantitative performance improvements, as classifiers utilizing these semantically enriched embeddings can more effectively differentiate between stress and non-stress content. For instance, in Figure 3(c), we observe that tokens associated with negative emotional states (which correlate with stress) form a cohesive cluster separate from neutral language, supporting our hypothesis that SCL effectively incorporates psychological knowledge about the relationship between emotions and stress into the embeddings.

In contrast, Figure 3(a), Figure 3(b), Figure 4(a), Figure 4(b), Figure 5(a) and Figure 5(b) shows that the baseline embeddings from SG_{PLM} without contrastive learning exhibit more homogeneous distribution, forming a single large cluster with less discernible separation between stress and non-stress semantic content. This comparative visualization directly illustrates the value added by our SCL-based approach in learning stress-specific features that mere integration of Skip-Gram with PLMs cannot achieve alone. We observed that some non-stress-related tokens appear within the stress-related cluster and vice versa, particularly for polysemous words that may appear in both stress and non-stress contexts. This finding suggests that while contrastive learning substantially enhances stress-specific representation learning, the lexicon-based identification of stress-indicative terms could benefit from more adaptive or data-driven strategies to further improve classification performance.

These visualizations provide qualitative evidence supporting our quantitative results, showing that StressSGCL successfully embeds domain-specific knowledge about stress into the word representation space, enabling the model to better distinguish stress-related content in social media posts across different platforms and contexts.

5 Conclusion and future work

In this paper, we have introduced StressSGCL, a novel stress detection approach that generates stress-specific embeddings capable of detecting stress in social media posts. It enhances the performance of classifiers in stress detection tasks by integrating the existing techniques of skip-gram

Table 2: Comparative performance evaluation of StressSGCL

Approaches	Accuracy			F_1			$MacroF_1$		
	DS ₁	DS ₂	DS ₃	DS ₁	DS ₂	DS ₃	DS ₁	DS ₂	DS ₃
StressSGCL _{BERT}	91.46	93.43	95.41	91.03	93.42	96.73	91.33	93.44	85.14
StressSGCL _{MentalBERT}	88.90	93.52	94.92	90.32	93.51	96.76	88.34	93.41	86.01
SSE _{BERT}	89.54	93.52	95.08	90.21	93.11	96.53	88.73	92.63	84.75
SSE _{MentalBERT}	80.36	92.52	95.16	90.21	92.54	96.32	88.43	92.41	84.38
TensiStrength	80.05	70.02	87.32	82.36	75.34	93.48	80.08	68.21	66.85
SG _{BERT}	87.61	92.14	94.44	89.48	92.47	97.53	87.50	92.11	85.10
SG _{MentalBERT}	88.10	92.27	94.48	90.15	92.42	97.36	88.63	92.21	84.63
BERT	75.23	84.30	89.02	77.31	83.58	95.28	74.61	84.31	71.21
MentalBERT	83.63	90.61	85.43	85.25	91.34	94.66	83.48	90.67	76.40

with transformers-based PLM using SCL to learn stress-specific embeddings. To capture stress-related emotions, StressSGCL employs TensiStrength lexicons. The TextCNN is used to classify whether the post belongs to the stress category by passing the stress-specific embeddings generated for the terms in the users' post. StressSGCL is evaluated on three publicly available benchmark datasets: Twitter Full, Reddit Title, and Reddit Combi. It surpasses the current methods and ablation-like baselines, thereby illustrating the efficacy of the proposed embeddings in comprehending the stress-related context. We intend to enhance StressSGCL in the future to identify chronic stress by examining a sequence of user posts. Furthermore, we intend to implement comparable methodologies to identify additional mental health conditions, including anxiety and depression.

References

- Bommasani, R.; Davis, K.; and Cardie, C. 2020. Interpreting pretrained contextualized representations via reductions to static embeddings. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 4758–4781.
- Chen, T.; Kornblith, S.; Norouzi, M.; and Hinton, G. 2020. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, 1597–1607.
- Chen, Y. 2015. *Convolutional neural network for sentence classification*. Master's thesis, University of Waterloo.
- 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.
- De Choudhury, M.; and De, S. 2014. Mental health discourse on reddit: Self-disclosure, social support, and anonymity. In *Proceedings of the international AAAI conference on web and social media*, volume 8, 71–80.
- De Choudhury, M.; Kiciman, E.; Dredze, M.; Coppersmith, G.; and Kumar, M. 2016. Discovering shifts to suicidal ideation from mental health content in social media. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, 2098–2110.
- Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Burstein, J.; Doran, C.; and Solorio, T., eds., *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 4171–4186. Minneapolis, Minnesota: Association for Computational Linguistics.
- Du, J.; Huang, J.; An, Y.; and Xu, W. 2018. The Relationship between stress and negative emotion: The Mediating role of rumination. *Clinical Research and Trials*, 4: 1–5.
- Gao, T.; Yao, X.; and Chen, D. 2021. Simcse: Simple contrastive learning of sentence embeddings. *arXiv preprint*.
- Ilias, L.; Mouzakitis, S.; and Askounis, D. 2023. Calibration of Transformer-Based Models for Identifying Stress and Depression in Social Media. *IEEE Transactions on Computational Social Systems*, 1–12.
- Jashinsky, J.; Burton, S. H.; Hanson, C. L.; West, J.; Giraud-Carrier, C.; Barnes, M. D.; and Argyle, T. 2014. Tracking suicide risk factors through Twitter in the US. *Crisis*.
- Ji, S.; Zhang, T.; Ansari, L.; Fu, J.; Tiwari, P.; and Cambria, E. 2022. MentalBERT: Publicly Available Pretrained Language Models for Mental Healthcare. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, 7184–7190.
- Khosla, P.; Teterwak, P.; Wang, C.; Sarna, A.; Tian, Y.; Isola, P.; Maschinot, A.; Liu, C.; and Krishnan, D. 2020. Supervised contrastive learning. *Advances in neural information processing systems*, 33: 18661–18673.
- Lin, H.; Jia, J.; Guo, Q.; Xue, Y.; Huang, J.; Cai, L.; and Feng, L. 2014a. Psychological stress detection from cross-media microblog data using deep sparse neural network. In *2014 IEEE international conference on multimedia and expo (ICME)*, 1–6. IEEE.
- Lin, H.; Jia, J.; Guo, Q.; Xue, Y.; Li, Q.; Huang, J.; Cai, L.; and Feng, L. 2014b. User-level psychological stress detection from social media using deep neural network. In *Proceedings of the 22nd ACM international conference on Multimedia*, 507–516.
- Lin, H.; Jia, J.; Nie, L.; Shen, G.; and Chua, T.-S. 2016. What Does Social Media Say about Your Stress?. In *IJCAI*, 3775–3781.
- Lin, H.; Jia, J.; Qiu, J.; Zhang, Y.; Shen, G.; Xie, L.; Tang, J.; Feng, L.; and Chua, T.-S. 2017. Detecting stress based on

social interactions in social networks. *IEEE Transactions on Knowledge and Data Engineering*, 29(9): 1820–1833.

Lokala, U.; Srivastava, A.; Dastidar, T. G.; Chakraborty, T.; Akhtar, M. S.; Panahiazar, M.; and Sheth, A. 2022. A computational approach to understand mental health from reddit: knowledge-aware multitask learning framework. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 16, 640–650.

Lu, H.; Frauendorfer, D.; Rabbi, M.; Mast, M. S.; Chittaranjan, G. T.; Campbell, A. T.; Gatica-Perez, D.; and Choudhury, T. 2012. Stresssense: Detecting stress in unconstrained acoustic environments using smartphones. In *Proceedings of the 2012 ACM conference on ubiquitous computing*, 351–360.

Mikolov, T.; Chen, K.; Corrado, G.; and Dean, J. 2013. Efficient estimation of word representations in vector space. *arXiv preprint*.

Oord, A. v. d.; Li, Y.; and Vinyals, O. 2018. Representation learning with contrastive predictive coding. *arXiv preprint*.

Poulin, C.; Shiner, B.; Thompson, P.; Vepstas, L.; Young-Xu, Y.; Goertzel, B.; Watts, B.; Flashman, L.; and McAllister, T. 2014. Predicting the risk of suicide by analyzing the text of clinical notes. *PLoS one*, 9.

Rajj, A.; Blitz, P.; Ali, A. A.; Fisk, S.; French, B.; Mitra, S.; Nakajima, M.; Nguyen, M. H.; Plarre, K.; Rahman, M.; et al. 2010. mstress: Supporting continuous collection of objective and subjective measures of psychosocial stress on mobile devices. *ACM Wireless Health 2010 San Diego, California USA*.

Rastogi, A.; Liu, Q.; and Cambria, E. 2022. Stress detection from social media articles: New dataset benchmark and analytical study. In *2022 International Joint Conference on Neural Networks (IJCNN)*, 1–8. IEEE.

Thelwall, M. 2017. TensiStrength: Stress and relaxation magnitude detection for social media texts. *Information Processing & Management*, 53(1): 106–121.

Turcan, E.; and McKeown, K. 2019. Dreddit: A reddit dataset for stress analysis in social media. *arXiv preprint*.

Turcan, E.; Muresan, S.; and McKeown, K. 2021. Emotion-infused models for explainable psychological stress detection. In *Proceedings of the 2021 conference of the North American Chapter of the Association for Computational Linguistics: human language technologies*, 2895–2909.

Vashishth, S.; Bhandari, M.; Yadav, P.; Rai, P.; Bhat-tacharyya, C.; and Talukdar, P. 2019. Incorporating Syntactic and Semantic Information in Word Embeddings using Graph Convolutional Networks. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 3308–3318.

Wang, X.; Zhang, H.; Cao, L.; Zeng, K.; Li, Q.; Li, N.; and Feng, L. 2023. Contrastive Learning of Stress-specific Word Embedding for Social Media based Stress Detection. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 5137–5149. Association for Computing Machinery.

Wang, Y.; Cui, L.; and Zhang, Y. 2021. Improving Skip-Gram Embeddings Using BERT. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29: 1318–1328.

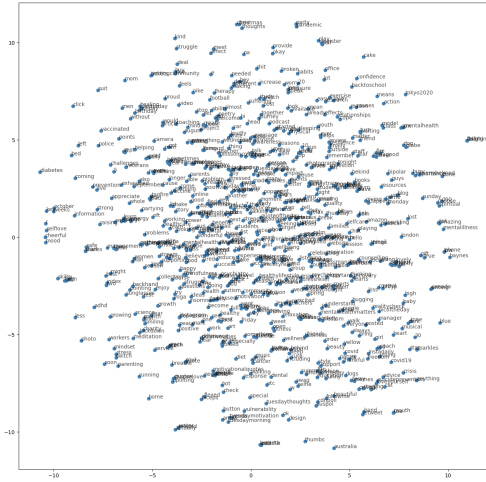
Wang, Y.; Wang, Z.; Li, C.; Zhang, Y.; and Wang, H. 2020. A multitask deep learning approach for user depression detection on sina weibo. *arXiv preprint*.

Yates, A.; Cohan, A.; and Goharian, N. 2017. Depression and self-harm risk assessment in online forums. *arXiv preprint*.

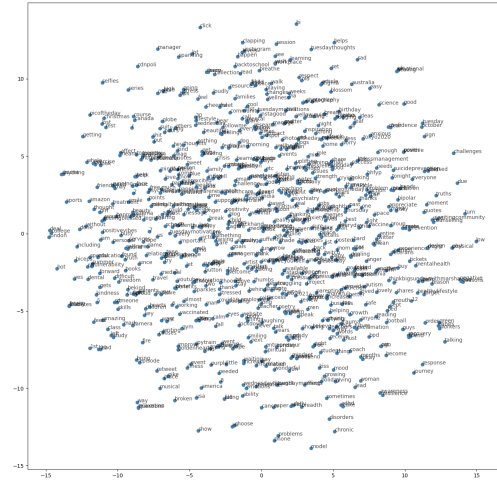
Paper Checklist

1. For most authors...
 - (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**
 - (b) Do your main claims in the abstract and introduction accurately reflect the paper’s contributions and scope? **Yes**
 - (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? **Yes**
 - (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? **Yes**
 - (e) Did you describe the limitations of your work? **Yes**
 - (f) Did you discuss any potential negative societal impacts of your work? **NA**
 - (g) Did you discuss any potential misuse of your work? **NA**
 - (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? **NA**
 - (i) Have you read the ethics review guidelines and ensured that your paper conforms to them? **Yes**
2. Additionally, if your study involves hypotheses testing...
 - (a) Did you clearly state the assumptions underlying all theoretical results? **NA**
 - (b) Have you provided justifications for all theoretical results? **NA**
 - (c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? **NA**
 - (d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? **NA**
 - (e) Did you address potential biases or limitations in your theoretical framework? **NA**
 - (f) Have you related your theoretical results to the existing literature in social science? **NA**
 - (g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? **NA**

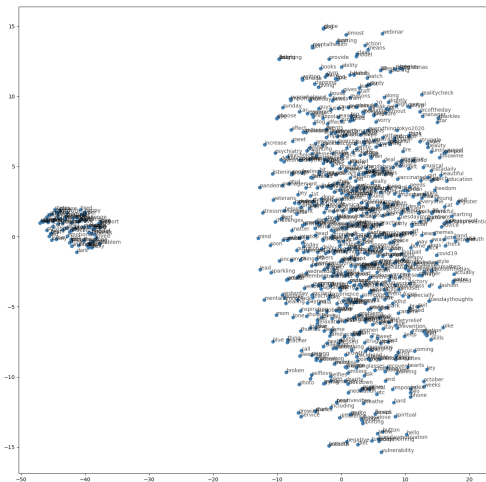
3. Additionally, if you are including theoretical proofs...
 - (a) Did you state the full set of assumptions of all theoretical results? **NA**
 - (b) Did you include complete proofs of all theoretical results? **NA**
4. Additionally, if you ran machine learning experiments...
 - (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? **Available on request.**
 - (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)? **No**
 - (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? **NA**
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? **NA**
 - (c) Did you include any new assets in the supplemental material or as a URL? **NA**
 - (d) Did you discuss whether and how consent was obtained from people whose data you’re using/curating? **NA**
 - (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? **NA**
 - (f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR? **NA**
 - (g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset? **NA**
6. Additionally, if you used crowdsourcing or conducted research with human subjects, **without compromising anonymity**...
 - (a) Did you include the full text of instructions given to participants and screenshots? **NA**
 - (b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? **NA**
 - (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? **NA**
 - (d) Did you discuss how data is stored, shared, and de-identified? **NA**



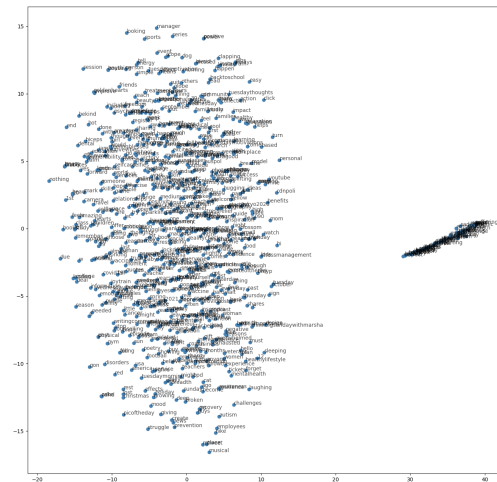
(a) SG_{BERT}



(b) SG_{MentalBERT}

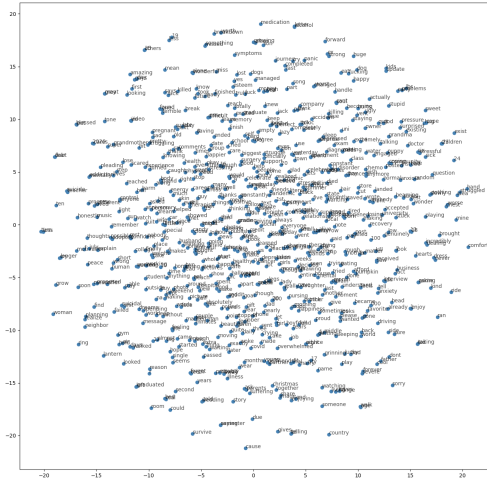


(c) StressSGCL_{BERT}

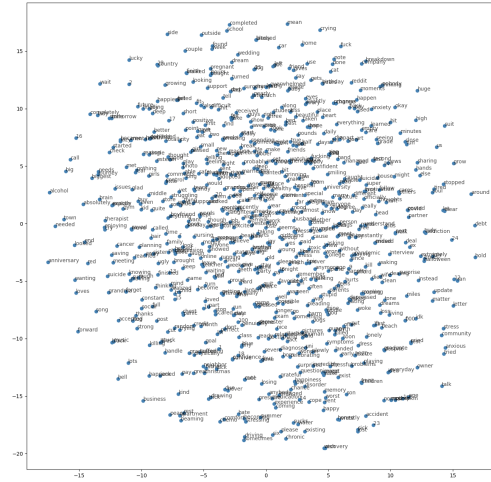


(d) StressSGCL_{MentalBERT}

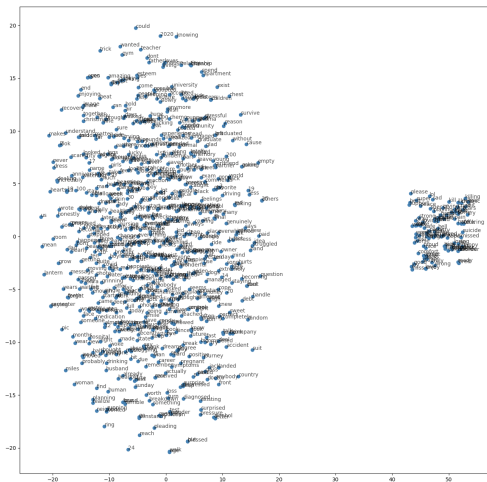
Figure 3: t-SNE visualizations of token embeddings for Twitter Full dataset.



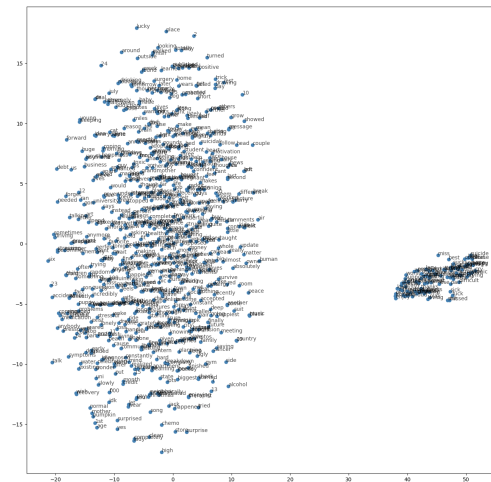
(a) SG_{BERT}



(b) $SG_{MentalBERT}$



(c) $StressSGCL_{BERT}$



(d) $StressSGCL_{MentalBERT}$

Figure 4: t-SNE visualizations of token embeddings for Reddit Title dataset.

