



Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Cabral, RC;Han, SC;Poon, J;Nenadic, G

Title:

MM-EMOG: Multi-Label Emotion Graph Representation for Mental Health Classification on Social Media †

Date:

2024-03-01

Citation:

Cabral, R. C., Han, S. C., Poon, J. & Nenadic, G. (2024). MM-EMOG: Multi-Label Emotion Graph Representation for Mental Health Classification on Social Media †. *Robotics*, 13 (3), <https://doi.org/10.3390/robotics13030053>.

Persistent Link:





<https://hdl.handle.net/11343/353068>

License:

[cc-by](#)

Article

# MM-EMOG: Multi-Label Emotion Graph Representation for Mental Health Classification on Social Media †

Rina Carines Cabral <sup>1,‡</sup> , Soyeon Caren Han <sup>1,2,\*</sup> , Josiah Poon <sup>1</sup>  and Goran Nenadic <sup>3</sup> 

<sup>1</sup> School of Computer Science, Faculty of Engineering, University of Sydney, Camperdown, NSW 2006, Australia; rina.cabral@sydney.edu.au (R.C.C.); josiah.poon@sydney.edu.au (J.P.)

<sup>2</sup> School of Computing and Information Systems, University of Melbourne, Parkville, VIC 3052, Australia

<sup>3</sup> School of Computer Science, University of Manchester, Manchester M13 9PL, UK; g.nenadic@manchester.ac.uk

\* Correspondence: caren.han@unimelb.edu.au or caren.han@sydney.edu.au

† Warning: This paper contains examples that are suicidal and depressive in nature.

‡ These authors contributed equally to this work.

**Abstract:** More than 80% of people who commit suicide disclose their intention to do so on social media. The main information we can use in social media is user-generated posts, since personal information is not always available. Identifying all possible emotions in a single textual post is crucial to detecting the user's mental state; however, human emotions are very complex, and a single text instance likely expresses multiple emotions. This paper proposes a new multi-label emotion graph representation for social media post-based mental health classification. We first construct a word–document graph tensor to describe emotion-based contextual representation using emotion lexicons. Then, it is trained by multi-label emotions and conducts a graph propagation for harmonising heterogeneous emotional information, and is applied to a textual graph mental health classification. We perform extensive experiments on three publicly available social media mental health classification datasets, and the results show clear improvements.

**Keywords:** mental health detection; emotion embedding; medical social media



**Citation:** Cabral, R.C.; Han, S.C.;

Poon, J.; Nenadic, G. MM-EMOG:

Multi-Label Emotion Graph

Representation for Mental Health

Classification on Social Media.

*Robotics* **2024**, *13*, 53. [https://doi.org/](https://doi.org/10.3390/robotics13030053)

[10.3390/robotics13030053](https://doi.org/10.3390/robotics13030053)

Academic Editor: Genci Capi

Received: 16 February 2024

Revised: 9 March 2024

Accepted: 15 March 2024

Published: 18 March 2024



**Copyright:** © 2024 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the Creative Commons

Attribution (CC BY) license ([https://creativecommons.org/licenses/by/](https://creativecommons.org/licenses/by/4.0/)

[https://creativecommons.org/licenses/by/](https://creativecommons.org/licenses/by/4.0/)

[4.0/](https://creativecommons.org/licenses/by/4.0/)).

## 1. Introduction

According to the World Health Organization [1] (WHO), it is revealed that a staggering majority of individuals who tragically succumb to suicide, surpassing 80%, choose to divulge their suicidal ideation and intentions on social media platforms. This trend underscores the profound impact and far-reaching implications that social media can have on mental health and well-being. The disclosure of such deeply personal and troubling thoughts on these digital platforms presents a unique opportunity for early intervention and detection of mental disorders, as well as potential suicidal tendencies. The significance of recognising and addressing these issues promptly extends beyond individual well-being to the broader spectrum of societal welfare, making early detection a vital component for fostering good governance. By understanding the intricate connection between social media expressions and mental health indicators, policymakers and healthcare professionals can proactively implement measures to provide timely support and intervention, thereby contributing to the overall enhancement of mental health outcomes on a societal level. In essence, the revelation of suicidal ideation on social media acts as a crucial signal for the imperative need to prioritise and enhance mental health surveillance and support systems for the greater well-being of communities.

One challenge of social media research, however, is the privacy and protection of one's identity. Users tend to prefer to be completely anonymous or to withhold personal details and confidential information, which is then impossible to observe. Accessing contextual components such as historical posts and user or post metadata information has also become

increasingly restrictive due to heightened data protocols from social media platforms, which complicates research reproducibility further. Due to this trend, the primary information that may be used for mental health detection from social media is user-generated posts. Our research focuses on detecting mental illnesses through the analysis of social media textual posts only with the question, ‘What would be the most important component from which we can identify the mental health condition using pure text from social media?’ The answer can be found in the WHO’s definition of mental disorder, stating that ‘A mental disorder is characterized by a clinically significant disturbance in an individual’s cognition, emotional regulation, or behaviour.’ [2]. The ideal setup for mental state detection via textual posts would identify all possible emotions and integrate those feelings and emotional statuses.

Recent studies incorporate emotions into mental health classification by fine-tuning pre-trained embeddings through a single emotion classification task [3,4]. Due to the complexity of human emotions, it is very likely that multiple emotions are expressed by a single textual post and that those emotions can be correlated. To represent emotions and their correlation with the text, we can consider two types of textual representation techniques: sequential text representation and graph-based text representation. While sequential text representation promotes capturing text features from local consecutive word sequences, graph-based text representation can attract widespread attention and successfully understand word and document relationships [5–7].

This paper proposes MM-EMOG, a new multi-label, graph-based emotion representation for mental health classification using user-generated social media posts. Note that we focus on the post-only-based mental health classification due to privacy issues and lack of personal information in social media. In order to achieve this aim, we first construct a word–document graph tensor to generate emotion-based contextual representation using emotion lexicons. It is then trained using multi-label emotions by conducting graph propagation and transformer backpropagation to harmonise heterogeneous emotional information. The trained multi-label emotion representation is applied with a textual graph mental health classification model.

The main contributions of this paper can be summarised as follows:

1. We propose a new multi-label emotion representation for mental health classification using only social media textual posts.
2. To our knowledge, no other studies have utilised Graph Convolutional Neural Network [8] (GCN) in a purely textual capacity for multi-label emotion representation and social media mental health classification tasks. We are the first to apply multi-label and graph-based textual emotion representation.
3. Our proposed model, MM-EMOG, achieved the highest performance on three publicly available social media mental health classification datasets.

## 2. Related Works

### 2.1. Social Media Mental Health Classification

Social media has opened up new opportunities for suicide ideation and mental health studies by creating a new way to access information-rich data not just of clinical patients but of a broader subset of the public outside clinical settings. Recent studies have focused on incorporating more social media components to capture as much available contextual information as possible. Among these are historical posts [4,9–18], conversation trees [19], social and interaction graphs [4,10,12,17,20], user and post metadata information [10,11], and images [10]. While more contextual sources may be ideal for assessing an individual’s mental health state, access to these data has become increasingly restrictive due to heightened data privacy concerns. It also complicates research reproducibility, since each study selects features based on what social media components are available to them. Because of this, our proposed system focuses on improving contextual information by incorporating emotions from the most basic social media component—a single textual post.

Emotions have been an area of interest for natural language processing (NLP) researchers because language has been one of the main avenues for communicating emo-

tions. We are particularly interested in emotions since mental health is deeply rooted in an individual's thoughts and feelings. Traditionally, frequency- or score-based emotion features [16,20–25] have been used to classify mental health states. For instance, Aragón et al. [21] proposed Bag-of-Sub-Emotions (BoSE), a histogram-based document emotion representation for the detection of depression and anorexia in online forums. Zogan et al. [25] implemented a feature selection process that includes positive, negative, and neutral emoji frequencies and valence, arousal, and dominance scores for depression detection. More recent studies have used machine and deep learning to fine-tune contextual embeddings using mental health classification as a downstream task [3,4,13,15,26]. In [26], an Emotion Understanding Network (EUN) is proposed where positive and negative word embeddings are learned separately by separate attention networks, which are later combined with a Semantic Understanding Network (SUN) for depression detection. Sawhney et al. [15] fine-tuned a pre-trained transformer on the EmoNet dataset [27] to extract the emotional spectrum of each post for suicide ideation detection in social media. However, these studies focus on learning a single emotion for a single word or an entire text. We recognise the complexity of human emotions, wherein a single word may express multiple emotions. Our proposed system integrates emotional context by harmonising heterogeneous emotions through a multi-label, corpus-based representation learning framework.

## 2.2. Graph Convolutional Networks

Graph Convolutional Networks (GCN) [8] have seen an increase in application in recent years due to their versatility in learning entity representations. TextGCN [5] adopts the GCN framework to learn representations for both words and documents without the use of external embeddings. Recent studies have improved upon this to learn multiple types of information through multi-edge [7] and multi-aspect [6] graphs. It has since been applied to long and short document classification [28,29], aspect-based sentiment analysis (ABSA) [30–32], and general text classification tasks [5,33]. Despite the incorporation of graph-based data, mental health classification in social media has been limited to user–user networks and user–post graphs [10,12,17,20]. To our knowledge, no other studies have utilised GCN in a purely textual capacity for these tasks. Our study will leverage the corpus-wide learning of TextGCN by thoroughly exhausting all co-occurrence relationships between words and documents. Our system utilises this corpus-based learning in a multi-label emotion framework to harmonise complex contextual and emotional information contained in mental-health-related textual posts.

## 3. MM-EMOG

### 3.1. MM-EMOG Construction

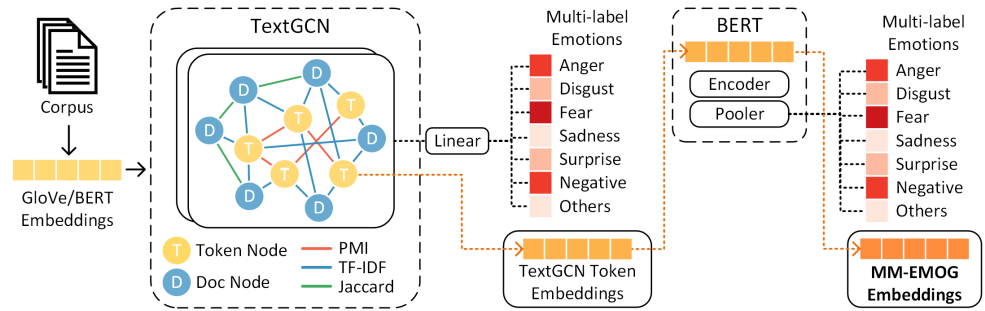
We adapt TextGCN [5] to learn the local and global emotional trend of mental health in social media via a graph-based structure  $G = (V, E, A)$ , where  $V$  is a set of word and document nodes,  $E$  is a set of word–word edges  $E_{w_i, w_j}$ , word–doc edges  $E_{w_i, d_j}$ , and doc–doc edges  $E_{d_i, d_j}$ , and  $A \in R^{N \times N}$  defines the weights of these associations. Figure 1 Step 1 shows the MM-EMOG architecture.

#### 3.1.1. Node Construction

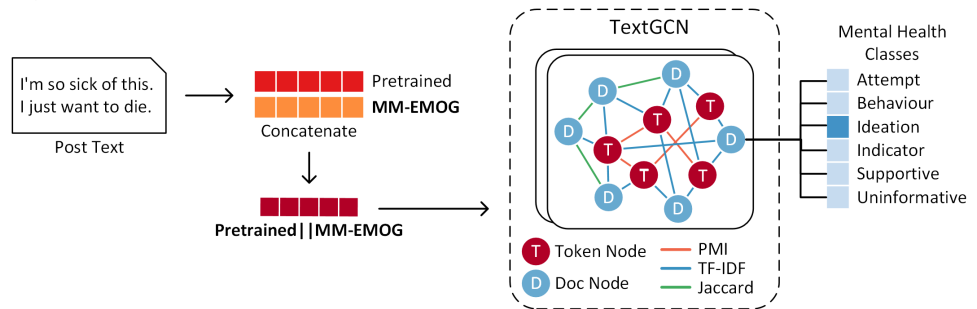
We first preprocess the post text in two steps. First, we further de-identified emails, usernames, and URLs by replacing them with reserved tokens. Second, we conduct emotion preservation by retaining emoticons and emojis to be contextualised as individual tokens. After preprocessing, we then create nodes by using each post as a document node and each token in the corpus as the word or token node. Token nodes are created either through (1) word split tokenisation or (2) wordpiece tokenisation using the pre-trained tokeniser from the Bidirectional Encoder Representations from Transformers (BERT) [34] implementation. For wordpiece tokens, we incorporate emoticons into the tokeniser vocabulary for emoticon preservation and only apply lowercasing without additional text preprocessing. For word split tokens, we employ a simple text cleaning process that re-

moves some punctuation and separates contractions. Stopwords are kept to retain negation words. Finally, we initialise word split token nodes using Global Vectors (GloVe) [35] as embeddings and use the average weight of all token nodes to represent the document node. For wordpiece tokens, we follow [36] and use contextualised BERT embeddings, where the learned vector for the [CLS] token is used to initialise each document node. For each unique token, all contextualized vector representation of the token is collected from the entire corpus. Minimum pooling is then applied to all the contextualized representations of the token to initialise the token node.

Step 1: Multi-label Emotion Embeddings



Step 2: Mental Health Classification



**Figure 1.** An overview of the proposed system showing the architectures for MM-EMOG (Step 1) and the mental health classification (Step 2). A textual graph is created to learn multi-emotion embeddings where nodes represent tokens and documents in the corpus while edges represent the relationships between them. The graph is passed to a two-layer GCN and a linear layer for a multi-label emotion classification task. After training, token node representations are extracted from the second GCN layer and used as initial weights for fine-tuning a pre-trained BERT model for the same task. After fine-tuning, the embeddings are extracted as the MM-EMOG embeddings for a graph-based mental health classification task.

3.1.2. Edge Construction

Inspired by [36], we leverage all types of co-occurrence relationships between tokens and documents using Pointwise Mutual Information (PMI) [37] for  $E_{w_i, w_j}$ , Term Frequency-Inverse Document Frequency (TF-IDF) for  $E_{w_i, d_j}$ , and Jaccard similarity for  $E_{d_i, d_j}$ . The final set of edges is  $E = \{E_{w_i, w_j}, E_{w_i, d_j}, E_{d_i, d_j}\}$ . Formally, the graph adjacency matrix is defined as:

$$A_{ij} = \begin{cases} PMI_{ij} & : i, j \text{ are words}; PMI > 0 \\ TF - IDF_{ij} & : i \text{ is word, } j \text{ is document} \\ Jaccard_{ij} & : i, j \text{ are documents} \\ 0 & : \text{otherwise} \end{cases}$$

### 3.2. MM-EMOG Learning

#### 3.2.1. Multi-Label Document Emotions

We first generate document-level, multi-label emotion classes to use as targets for generating emotion-rich contextual representations using emotion lexicons (Section 4.2) containing word–emotion associations. Note that we refer to both emotion and sentiment as “emotion” for the rest of the paper. Assume a document with words  $W = \{w_1, \dots, w_p\}$  where  $p$  is the number of unique words in the document and a lexicon containing terms  $K = \{k_1, \dots, k_q\}$  where  $q$  is the number of lexicon terms. Each lexicon term  $k_j$  is associated with one or more emotions  $EM = \{em_1, \dots, em_r\}$  where  $r$  is the number of emotion classes in the lexicon. We group positive emotions into “other” as higher-risk classes are more affected by negative emotions. When  $w_i = k_j$ , we extract the emotions  $EM_{k_j}$  associated with  $w_i$ . The final multi-label emotion class for the document is the union of all emotions associated with all of the words in the document  $EM_d = \{EM_{w_1} \cup EM_{w_2} \cup \dots \cup EM_{w_p}\}$ .

#### 3.2.2. Multi-Label Emotion Training

To incorporate complex emotions into contextual embeddings, we conduct a multi-label emotion classification task by passing the node representations  $V$  and the adjacency matrix  $A$  to a two-layer GCN followed by a linear layer with an output dimension of  $r$  emotions. The second layer of the GCN has a dimension of  $s = 768$  to follow popular pre-trained embeddings such as BERT [34]. Graph propagation takes the input and maps each instance to multiple emotions. In particular, the first GCN layer takes the input graph  $G$  to learn an input-to-hidden weight matrix  $W^{(0)} \in R^{d \times H}$  [8], where  $d$  is the dimension of node representations and  $H$  is the hidden dimension size. This weight matrix is subsequently passed to a second GCN layer to produce a hidden-to-hidden weight matrix  $W^{(1)} \in R^{H \times s}$ . The final linear layer learns the hidden-to-output weights  $W^{out} \in R^{s \times r}$ . ReLu is used with binary cross-entropy loss for multi-label learning. Backpropagation updates the initial representations to incorporate emotional information during model training. The learned token node representations  $W^{(1)}$  from the second GCN layer are extracted and used as the initial weights for fine-tuning a pre-trained BERT on the same multi-label emotion classification task using the last hidden layer of the [CLS] token. Similarly, we use the binary cross entropy loss function. The learned weights are extracted as multi-emotion contextual representations, MM-EMOG EmoWord (EW) or EmoWordPiece (EWP) embeddings.

### 3.3. Mental Health Post Classification

We evaluate MM-EMOG through a mental health post-based classification task (Figure 1 Step 2). We recognise the benefits of a user-based classification; however, post-based classification is more scalable and would not require other social media components such as historical posts. Similar to Step 1, we leverage the corpus-wide co-occurrence information from TextGCN using the same graph construction method. For token node representations, we concatenate BERT and MM-EMOG embeddings, while for document node representations, the average of all token representations within each document is used. Finally, the graph is passed to a two-layer GCN for mental health post classification following a similar graph propagation as the MM-EMOG training. However, for mental health post classification, the second GCN layer learns a hidden-to-output weight matrix  $W^{mh} \in R^{H \times c}$ , where  $c$  is the number of mental health classes. Categorical cross-entropy is used for single-label classification.

## 4. Experimental Setup

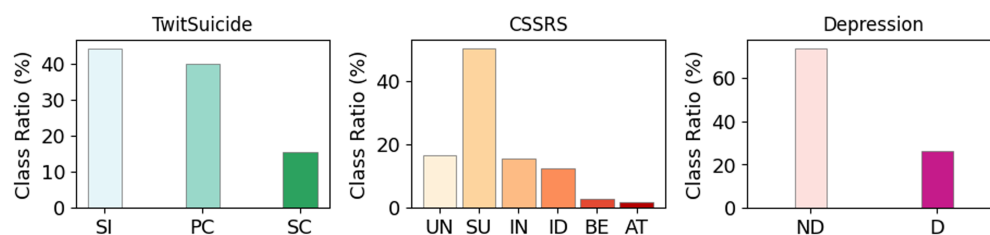
### 4.1. Datasets

We use three publicly available post-based datasets to evaluate MM-EMOG. Table 1 summarises the statistics, while Figure 2 shows the distribution of classes for each dataset. **TwitSuicide** (data available upon request) [38] is a dataset of Twitter posts gathered by searching suicide-related terms and annotated by one psychologist and two computer scientists based on three risk levels outlined by [39]: *Strongly Concerning* (SC; 15.61%), *Possibly Concerning* (PC;

40.00%), and *Safe to Ignore* (SI; 44.39%). The Reddit C-SSRS Dataset (**CSSRS**) (<https://github.com/AmanuelF/Suicide-Risk-Assessment-using-Reddit>) [22] is acquired from 15 mental-health-related subreddits annotated by four clinical psychiatrists based on the Columbia-Suicide Severity Rating Scale [40]. We use the post-level annotations with six classes: *Actual Attempt* (AT; 1.83%), *Suicidal Behavior* (BE; 2.87%), *Suicidal Ideation* (ID; 12.57%), *Suicidal Indicator* (IN; 15.67%), *Supportive* (SU; 50.45%), and *Uninformative* (UN; 16.60%). We note that the dataset contains medical entity normalised posts as detailed in the original authors' paper. The Twitter Depression Dataset (**Depression**) (<https://github.com/swcwang/depression-detection>) (accessed on 17 March 2024) is used as a basis for the practice dataset for CLPsych 2021 [41]. Tweets are collected using depression-related hashtags, stripped off of hashtags, and annotated using binary classes *Depression* (D; 26.34%) and *Non-Depression* (ND; 73.66%).

**Table 1.** Dataset statistics.

	TwitSuicide	CSSRS	Depression
Platform	Twitter	Reddit	Twitter
Total Posts	660	2680	3200
Total Users	645	375	-
Number of Classes	3	6	2
Evaluation Method	10 CV	5 CV	80/20
Length	13–147	2–6221	6–374
Average Length	90.32	451.67	90.08
Word Count	3–31	1–1051	1–77
Average Word Count	16.85	85.51	17.43



**Figure 2.** Class distribution for each dataset. For TwitSuicide (**left**), SI: Safe to Ignore, PC: Possibly Concerning, SC: Strongly Concerning. For CSSRS (**center**), UN: Uninformative, SU: Supportive, IN: Indicator, ID: Ideation, BE: Behaviour and AT: Attempt. For Depression (**right**), ND: Non-depression and D: Depression.

#### 4.2. Emotion Lexicons

To create emotion-rich contextual embeddings that capture the complex emotions within mental health-related posts, we use three widely used emotion lexicons that associate one or more emotions or sentiments to words or concepts. Table 2 enumerates the emotion types for each lexicon. The NRC Emotion Lexicon (**EmoLex**) (<https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm>) (accessed on 17 March 2024) [42] is a crowdsourced word–emotion and word–polarity pairings. The lexicon contains 6453 terms matched to at least one of two sentiments or eight emotions. The NRC Twitter Emotion Corpus (**TEC**) (<http://saifmohammad.com/WebPages/lexicons.html>) (accessed on 17 March 2024) [43] is an automatically created lexicon using emotion hashtags from Twitter. Word co-occurrence scores determine the word–emotion association. We apply a threshold of at least 0.5 to remove weakly associated pairs. A total of 16,862 terms, including hashtags, emoticons, common stop words, proper names, and numerical figures, are associated with at least one of eight emotions. **SenticNet** (<https://sentic.net/downloads/>) (accessed on 17 March 2024) [44] is a concept-level knowledge base created through commonsense knowledge graphs. We use SenticNet7, which generates symbolic representations through

subsymbolic techniques. A total of 149,673 concepts, including emoticons and emojis, are associated with one sentiment and two of 24 fine-grained emotions. We simplify these to eight primary emotions by grouping them based on their positive and negative intensity levels. Furthermore, for simplicity, we only utilise one-word concepts.

**Table 2.** Emotion types for each lexicon.

Lexicon	Description	Emotion Types
EmoLex [42]	Crowdsourced word–emotion and word–polarity pairings	Anger, Anticipation *, Disgust, Fear, Joy *, Sadness, Surprise, Trust *, Positive *, Negative
TEC [43]	Automatic annotation from emotion-related hashtags on Twitter	Anger, Anticipation *, Disgust, Fear, Joy *, Sadness, Surprise, Trust *
SenticNet [44]	Concepts from common sense knowledge graphs associated with emotions through similarity prediction	Anger, Calmness *, Disgust, Eagerness *, Fear, Joy *, Pleasantness *, Sadness, Positive *, Negative

\* Combined into “other”.

#### 4.3. Baselines and Metrics

We evaluate the performance of our proposed system against four transformer-based pre-trained language models (PLM). **BERT** [34] was trained on next sentence prediction and masked language modelling that generated state-of-the-art performance on multiple NLP-related tasks. **RoBERTa** [45] replicated the BERT training setup but used a dynamic masking pattern and removed the next sentence prediction objective. **MentalBERT** and **MentalRoBERTa** [46] followed the training procedures of BERT and RoBERTa, but used data from different mental health subreddits, including *r/SuicideWatch*, *r/Anxiety*, *r/bipolar*, *r/mentalillness*, and *r/mentalhealth*.

All baseline models are trained for 15 epochs with a 0.0001 learning rate, a maximum length of 256, and a batch size of 8. Other hyperparameters are left to the default values set by the HuggingFace (<https://huggingface.co/> (accessed on 17 March 2024)) library. Due to class imbalance, we evaluate our models’ overall performance using accuracy and weighted F1 (F1w) scores and provide a detailed breakdown for each class using class F1 scores.

#### 4.4. Implementation Details

We model MM-EMOG using a train/validation split of 90:10 and train for 200 epochs with a 10-epoch early stop using the Adam optimiser for both TextGCN and BERT phases. The TextGCN phase uses the following hyperparameters: 200 hidden dimension size, 0.5 dropout, 0.02 learning rate, and 2 GCN layers. The BERT phase uses 0.5 dropout,  $1 \times 10^{-5}$  learning rate, and 256 maximum length. Batch size is set to 64, 32, and 16 for the TwitSuicide, CSSRS, and Depression datasets, respectively.

For the classification task, we use established evaluation setups following previous studies for each dataset: cross-validation (CV) with 10- and 5-fold setups for TwitSuicide and CSSRS, respectively, and a 80:20 train/test split for Depression. All classification models are tuned using Optuna (<https://github.com/optuna/optuna> (accessed on 17 March 2024)), a hyperparameter optimization framework, at a 90:10 validation split. The hyperparameters tuned are the number of hidden layers  $L = \{2, 3, 4, 5\}$ , hidden layer dimension  $H = \{100, 200, 300, 400, 500\}$ , dropout  $dr = \{0.01, 0.05, 0.1, 0.5\}$ , learning rate  $lr = \{0.01, 0.02, 0.03, 0.04, 0.05\}$ , and weight decay  $wd = \{0, 0.005, 0.05\}$ . Results are reported in Table 3 based on an average of 10 independent runs on Google Colab GPU-hosted runtimes. Best-found hyperparameters for each model are presented on Appendix A.

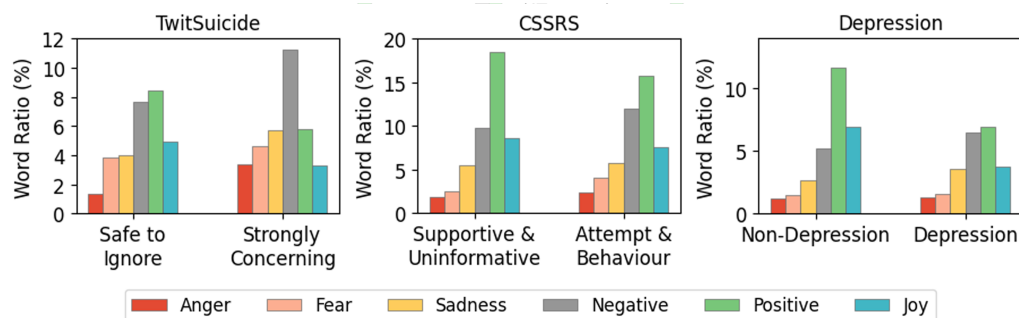
**Table 3.** The overall results of our mental health classification model using MM-EMOG representations concatenated with BERT and over different emotion-based lexicons. The best scores are **bold faced**; the second best are underlined. Class-based scores are shown for the most and least concerning classes for each dataset. For TwitSuicide: Strongly Concerning (SC) and Safe to Ignore (SI). For CSSRS: weighted average of Attempt, Behaviour, and Ideation (A,B,I) and Uninformative (UN). For Depression: Depression (D) and Non-Depression (ND).

TwitSuicide	Acc	F1w	Class F1-Scores	
			(SC)	(SI)
BERT	55.15	54.25	33.96	61.49
RoBERTa	45.00	38.86	00.00	60.43
MentalBERT	63.33	63.29	<u>48.00</u>	71.23
MentalRoBERTa	45.75	44.02	24.46	53.22
Ours (EW2-EmoLex)	67.97	65.26	28.06	75.96
Ours (EW2-TEC)	<b>71.86</b>	<b>71.03</b>	<b>52.64</b>	<b>78.03</b>
Ours (EW2-SenticNet)	<u>70.12</u>	<u>68.80</u>	44.09	<u>76.84</u>
CSSRS	Acc	F1w	(A,B,I)	(UN)
BERT	53.02	44.38	16.75	22.59
RoBERTa	28.66	25.86	00.00	23.38
MentalBERT	51.75	50.02	28.84	35.16
MentalRoBERTa	36.04	30.92	00.00	21.75
Ours (EWP1-EmoLex)	<b>73.07</b>	<b>70.79</b>	<b>43.82</b>	<b>72.71</b>
Ours (EWP1-TEC)	<u>72.34</u>	<u>69.79</u>	<u>41.54</u>	<u>72.09</u>
Ours (EWP1-SenticNet)	70.07	67.41	37.86	71.14
Depression	Acc	F1w	(D)	(ND)
BERT	73.59	62.40	00.00	84.79
RoBERTa	73.59	62.40	00.00	84.79
MentalBERT	73.59	62.40	00.00	84.79
MentalRoBERTa	73.59	62.40	00.00	84.79
Ours (EWP2-EmoLex)	77.56	<b>76.61</b>	<b>52.31</b>	85.33
Ours (EWP2-TEC)	<u>77.64</u>	<b>76.61</b>	<u>49.40</u>	<u>85.61</u>
Ours (EWP2-SenticNet)	<b>78.16</b>	<u>76.20</u>	48.51	<b>86.13</b>

EW: word split; EWP: word piece; 1: simple cleaning; 2: added de-identification and emoticon preservation.

## 5. Emotion Analysis

Our main goal for this study is to create contextualised representations incorporating heterogeneous emotions associated with mental-health-related text. This goal depends heavily on learning emotions associated with words. With this, we evaluate the use of emotion lexicons to create multi-label emotion classes for our study. After matching post words to lexicon emotions, we find an increase of negative emotions from the least to the most concerning classes. In contrast, a negative trend emerges for the positive emotions (Figure 3). This trend demonstrates how social media posts contain emotional markers consistent with different levels of suicide ideation and depression. The heterogeneity of these emotions motivates the use of a multi-label approach in learning emotional contextual representations for mental health classification.



**Figure 3.** Emotion label distribution using the SenticNet lexicon on the three benchmark datasets.

## 6. Results

### 6.1. Overall Performance

We evaluate the MM-EMOG representations through a mental health classification task using a graph-based model. Table 3 shows results from our proposed system against pre-trained language model baselines. Due to the small percentages of the most concerning classes for the CSSRS dataset, we report a combined weighted F1-score for *Attempt* (AT), *Behaviour* (BE), and *Ideation* (ID) classes (Section 4.1). Overall, our system outperforms all the baselines with an 8%, 21%, and 14% improvement for TwitSuicide, CSSRS, and Depression, respectively. Moreover, a notable increase in performance over the most concerning classes shows that, through multi-label contextual emotion representation learning, MM-EMOG can capture emotional intricacies where heightened negative emotions are present. We note that due to the severe binary class imbalance of 74:26, all the baselines for the Depression dataset are only predicting the majority class. Without using class weights or balancing methods, using MM-EMOG produces better performance.

### 6.2. Ablation Results

To analyse what lexical components are beneficial for learning contextual emotional representations, we compare different embeddings based on the lexicon used to train them in Table 3. Twitter-based datasets achieve better performance when trained with TEC and SenticNet, which both include hashtags, emoticons, or emojis that are more frequently used on Twitter than on Reddit. These results imply the importance of including these components in learning emotion representations for social media.

We also compare the effect of different tokenisation methods and further de-identification and emoticon preservation (Section 3.1). Table 4 shows that Twitter-based datasets perform better for de-identified and emoticon-preserved setups, possibly due to the frequent use of usernames, URLs, emoticons, and emojis on the platform. De-identification reduces noise during model training, while preserving emoticons as separate tokens contextualises them in the same way that words are contextualised with different meanings. Comparing tokenisation setups, both the CSSRS and the Depression datasets achieve better performance when wordpiece tokenised, while a simple word split is better for TwitSuicide. We note that during graph construction using the word split setup, TwitSuicide's vocabulary size is only 330, while Depression and CSSRS datasets have 1178 and 2673, respectively. The smaller vocabulary graph of TwitSuicide might have allowed it to perform better on a word split setup. Longer and larger datasets benefit more from wordpiece tokenisation because of the deconstruction of out-of-vocabulary words.

Finally, we compared concatenating MM-EMOG embeddings with BERT or MentalBERT embeddings for the mental health classification task in Table 5. Using the best setup and lexicon for each document from the previous tests, BERT performs slightly better than MentalBERT, despite being trained on mental-health-related data; thus, we retain the use of BERT embeddings for the rest of the experiments.

**Table 4.** Ablation study comparing the accuracy achieved using different text preprocessing setups. EW: word split; EWP: word piece; 1: simple cleaning; 2: added de-identification and emoticon preservation. The best scores are **bold faced**; the second best are underlined.

Dataset (Lexicon)	EW1		EW2		EWP1		EWP2	
	Acc	F1w	Acc	F1w	Acc	F1w	Acc	F1w
TwitSuicide (TEC)	<u>69.24</u>	<u>67.01</u>	<b>71.86</b>	<b>71.03</b>	67.52	64.81	68.09	65.73
CSSRS (EmoLex)	70.30	66.99	72.33	69.81	<b>73.07</b>	<b>70.79</b>	<u>72.59</u>	<u>70.28</u>
Depression (SenticNet)	76.06	66.46	<u>77.45</u>	66.22	77.27	<u>68.83</u>	<b>78.16</b>	<b>76.20</b>

**Table 5.** Comparison of BERT and MentalBERT as the pretrained embedding concatenated with MM-EMOG for mental health classification. EW: word split; EWP: word piece; 1: simple cleaning; 2: added de-identification and emoticon preservation. The best scores are **bold faced**.

Dataset	Setup	Lexicon	Embedding	Acc	F1w
TwitSuicide	EW2	TEC	BERT	<b>71.86</b>	<b>71.03</b>
			MentalBERT	70.55	69.44
CSSRS	EWP1	EmoLex	BERT	<b>73.07</b>	<b>70.79</b>
			MentalBERT	72.02	69.47
Depression	EWP2	SenticNet	BERT	<b>78.16</b>	<b>76.20</b>
			MentalBERT	77.48	<b>76.20</b>

### 6.3. Case Studies

We further evaluate MM-EMOG with a qualitative assessment of the produced predictions. In Table 6, each sample is compared to the prediction of the two best-performing baseline models, BERT and MentalBERT. We note that for the *Ideation* (ID) class of CSSRS, our system distinguishes between simultaneous expression of support and ideation. Expressions of empathy such as “I know what you mean” and “I feel the same way” are frequently expressed in the *Supportive* (SU) class; however, these are directed toward situations that trigger negative emotions, such as having no one to talk to or being in an unpleasant environment. For the ID class, empathy is expressed towards hopelessness and self-harm. MM-EMOG captures emotional context that differentiates these better.

**Table 6.** Qualitative comparison of MM-EMOG predictions over the two best performing baseline models: BERT and MentalBERT. Parts of the examples are masked with \*\*\* to prevent a reverse search for each post. The **bold text** shows correct predictions by the model.

Example	Actual	Ours	BERT	MentalBERT
<b>TwitSuicide</b>				
i’m SO fucking tired i want to die. *** adrenal exhaustion *** since surgery, I have not been well ***	SC	<b>SC</b>	PC	PC
*** tired, *** foot hurts *** do not want to be here	PC	<b>PC</b>	SC	SC
*** victim of a failed suicide attempt *** I dont wet-shave my neck. Ouch	SI	<b>SI</b>	PC	SC
<b>CSSRS</b>				
Aannnnnnd I failed... again. *** pills *** stomach Muscle cramp and Common cold chills...	AT	<b>AT</b>	SU	IN
*** VA hospital for three months *** awesome.	BE	<b>BE</b>	SU	BE

Table 6. Cont.

Example	Actual	Ours	BERT	MentalBERT
I know what you mean. I think about blowing my brains *** the immensely sweet relief *** constant Anxiety and Fear no longer exist. All of my issues will disappear, and thats all that matters. Why is suicide bad, again? *** why should I continue? ***	ID	ID	SU	SU
*** Im still sad that I had to go trough my life, sometimes bit angry to fate, *** nothing to show of my life. *** no longer bitter and *** that I was/am bad and deserved this. ***	IN	IN	SU	ID
*** you didnt study the right way :) Things change *** so dont give up! I thought I wouldnt make it *** but then I changed majors ***	SU	SU	IN	UN
*** dressed in some of my finer casual *** made myself some coffee. *** today is better ***	UN	UN	SU	AT
<b>Depression</b>				
*** scares get re opened *** pooring salt in them. I hate this feeling. *** pain im in again	D	D	ND	ND
*** so revolting, yet so irresistible *** I must have it	ND	ND	ND	ND

## 7. Ethical Considerations

While our work is mainly at a foundational research stage and not yet for production and deployment, we recognise that mental health classification using social media may be used to profile and disadvantage people with mental health issues in certain situations, such as employment and housing applications. However, we aim for the safeguarded use of any future healthcare application borne from this research, primarily for early detection and prevention of extreme outcomes of mental illnesses, such as self-harm and suicide. Two possible future applications are (1) for individual patient monitoring at the hands of mental health experts with proper patient consent or (2) for population-level monitoring for better mental health resource planning.

The use of publicly available data from social media comes with inherent risks, which we attempt to mitigate when conducting the study. First, we further de-identified each post in each dataset by replacing usernames, hyperlinks, and email addresses. Furthermore, we made it a point to mask published examples to prevent reverse searches leading back to the poster's account. We discuss other limitations of the study on Appendix B.

## 8. Conclusions

Mental Illness Detection through individual social media posts is a challenging task due to limited information. Since mental health is deeply rooted in emotions, identifying all possible emotions within the text is crucial to enrich contextual representations further. We introduced MM-EMOG (Multi-label Mental Health Emotion Graph) representations, which contextualise and harmonise complex heterogeneous emotions through a corpus-based, multi-label learning framework. MM-EMOG representations are learned through a multi-label emotion classification task using GCN to leverage the global and local relationship of words and documents, followed by BERT for enhanced semantic contextualisation of each token embedding. We evaluate MM-EMOG through a graph-based mental health classification task. Our results show that MM-EMOG successfully outperforms baselines in three social media mental health datasets with notable improvements over the most concerning classes regardless of the lexicon used for pre-training. Tokenisation methods and further de-identification and emoticon preservation affect the datasets differently due to the different characteristics of the datasets. In the future, we aim to release a pre-

trained MM-EMOG model with generalised emotion representations for mental health downstream tasks.

**Author Contributions:** Conceptualization, S.C.H.; methodology, R.C.C. and S.C.H.; validation, R.C.C.; formal analysis, R.C.C., S.C.H. and G.N.; investigation, R.C.C.; writing—original draft preparation, R.C.C.; writing—review and editing, S.C.H., J.P. and G.N.; visualization, R.C.C.; supervision, S.C.H. and J.P.; project administration, S.C.H.; funding acquisition, S.C.H. and J.P. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by funding from Google Award for inclusion research program (Project Name: Multimodal Mental Illness Detection and Explanation).

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of the University of Sydney (Project Number: 2019/741).

**Data Availability Statement:** We make the code available at [https://github.com/adlnlp/mm\\_emog](https://github.com/adlnlp/mm_emog). The TwitSuicide dataset is available upon request from the corresponding author. The CSSRS dataset may be downloaded from <https://github.com/AmanuelF/Suicide-Risk-Assessment-using-Reddit> and Depression dataset from <https://github.com/swcwang/depression-detection>. EmoLex and TEC lexicons may be obtained from <https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm> and <https://saifmohammad.com/WebPages/lexicons.html>. SenticNet is available at <https://sentic.net/downloads/>. All datasets and lexicons were accessed on 17 March 2024.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

WHO	World Health Organization
MM-EMOG	Multi-label Mental Health Emotion Graph representations
NLP	Natural Language Processing
GCN	Graph Convolutional Neural Network
BERT	Bidirectional Encoder Representations from Transformers
RoBERTa	Robustly Optimized BERT Pretraining Approach
GloVe	Global Vectors for Word Representation
PMI	Pointwise Mutual Information
TF-IDF	Term Frequency-Inverse Document Frequency
ReLU	Rectified Linear Unit
CV	Cross Validation
SC	Strongly Concerning
PC	Possibly Concerning
SI	Safe to Ignore
AT	Actual Attempt
BE	Suicidal Behaviour
ID	Suicidal Ideation
IN	Suicidal Indicator
SU	Supportive
UN	Uninformative
D	Depression
ND	Non-Depression
Acc	Accuracy
F1w	Weighted F1-score
EW	EmoWord representation
EWP	EmoWordPiece representation

## Appendix A. Hyperparameter Search

We utilised Optuna (<https://github.com/optuna/optuna> (accessed on 17 March 2024)) to search for optimal hyperparameters for the mental health classification task. Each

model setup was searched separately for 50 trials, maximising accuracy using a 90:10 split of the whole dataset for cross-validated datasets or of the training set for datasets with defined splits. We searched for the following hyperparameters: number of hidden layers  $L = \{2, 3, 4, 5\}$ , hidden layer dimension  $H = \{100, 200, 300, 400, 500\}$ , dropout  $dr = \{0.01, 0.05, 0.1, 0.5\}$ , learning rate  $lr = \{0.01, 0.02, 0.03, 0.04, 0.05\}$ , and weight decay  $wd = \{0, 0.005, 0.05\}$ . Best-found values are summarised in Table A1.

**Table A1.** Best -found hyperparameters for each dataset using all lexicons and all preprocessing setups. Emo: EmoLex; Sen: SenticNet.

	TwitSuicide			CSSRS			Depression		
	Emo	TEC	Sen	Emo	TEC	Sen	Emo	TEC	Sen
<b>EW1</b>									
dropout	0.5	0.01	0.5	0.01	0.1	0.05	0.01	0.1	0.05
num layers	4	2	2	2	2	2	2	2	2
num hidden	200	400	400	300	200	500	200	200	200
learning rate	0.01	0.03	0.04	0.05	0.03	0.03	0.05	0.02	0.05
weight decay	0	0	0	0	0	0	0	0	0
<b>EW2</b>									
dropout	0.5	0.01	0.01	0.01	0.01	0.05	0.01	0.5	0.05
num layers	2	2	2	2	2	2	2	2	2
num hidden	200	200	400	300	400	200	200	200	200
learning rate	0.02	0.05	0.01	0.03	0.05	0.04	0.05	0.03	0.01
weight decay	0	0	0	0	0	0	0	0	0
<b>EWP1</b>									
dropout	0.5	0.01	0.1	0.01	0.1	0.5	0.1	0.5	0.1
num layers	2	2	2	2	2	2	2	2	2
num hidden	100	100	200	200	200	400	200	200	200
learning rate	0.05	0.01	0.04	0.04	0.04	0.05	0.01	0.02	0.05
weight decay	0	0	0	0	0	0	0	0	0
<b>EWP2</b>									
dropout	0.5	0.5	0.1	0.01	0.5	0.05	0.05	0.1	0.01
num layers	2	2	5	2	2	2	2	2	2
num hidden	200	500	300	200	500	200	200	200	200
learning rate	0.02	0.04	0.04	0.04	0.04	0.05	0.04	0.05	0.02
weight decay	0	0	0.005	0	0	0	0	0	0

EW: word split; EWP: word piece; 1: simple cleaning; 2: added de-identification and emoticon preservation.

## Appendix B. Limitations

We acknowledge three limitations of our study. First, we use mainly English-based datasets, lexicons, and baseline models. Low-resource languages were not explored in this study, but it is an open direction for the future. We also note that despite being marked as English, some posts may contain a mix of different languages. Second, the computational resource needed for building and training graph networks grows exponentially with the length and size of the datasets. We are limited by the resources available to us, which only allow a maximum of 256 words from each post. Lastly, there are not enough publicly available state-of-the-art models for single post-only, text-based mental health classification. Thus, we provide baselines based on widely used pre-trained language models.

## References

1. World Health Organization. One in 100 Deaths is by Suicide. *World Health Organization News Release*, 17 June 2021. Available online: <https://www.who.int/news/item/17-06-2021-one-in-100-deaths-is-by-suicide> (accessed on 17 March 2024).
2. World Health Organization. Mental Disorders. 2022. Available online: <https://www.who.int/news-room/fact-sheets/detail/mental-disorders> (accessed on 17 March 2024).

3. Lara, J.S.; Aragón, M.E.; González, F.A.; Montes-y Gómez, M. Deep Bag-of-Sub-Emotions for Depression Detection in Social Media. In Proceedings of the Text, Speech, and Dialogue: 24th International Conference, TSD 2021, Olomouc, Czech Republic, 6–9 September 2021; Springer: Cham, Switzerland, 2021; pp. 60–72. [\[CrossRef\]](#)
4. Sawhney, R.; Joshi, H.; Shah, R.R.; Flek, L. Suicide Ideation Detection via Social and Temporal User Representations using Hyperbolic Learning. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Virtual, 6–11 June 2021; pp. 2176–2190. [\[CrossRef\]](#)
5. Yao, L.; Mao, C.; Luo, Y. Graph Convolutional Networks for Text Classification. *Proc. AAAI Conf. Artif. Intell.* **2019**, *33*, 7370–7377. [\[CrossRef\]](#)
6. Liu, X.; You, X.; Zhang, X.; Wu, J.; Lv, P. Tensor Graph Convolutional Networks for Text Classification. *Proc. AAAI Conf. Artif. Intell.* **2020**, *34*, 8409–8416. [\[CrossRef\]](#)
7. Wang, K.; Han, S.C.; Long, S.; Poon, J. ME-GCN: Multi-dimensional Edge-Embedded Graph Convolutional Networks for Semi-supervised Text Classification. *arXiv* **2022**, arXiv:2204.04618. [\[CrossRef\]](#)
8. Kipf, T.N.; Welling, M. Semi-supervised classification with graph convolutional networks. *arXiv* **2016**, arXiv:1609.02907.
9. Buddhitha, P.; Inkpen, D. Multi-task learning to detect suicide ideation and mental disorders among social media users. *Front. Res. Metr. Anal.* **2023**, *8*, 1152535. [\[CrossRef\]](#)
10. Cao, L.; Zhang, H.; Feng, L. Building and Using Personal Knowledge Graph to Improve Suicidal Ideation Detection on Social Media. *IEEE Trans. Multimed.* **2022**, *24*, 87–102. [\[CrossRef\]](#)
11. Cao, L.; Zhang, H.; Feng, L.; Wei, Z.; Wang, X.; Li, N.; He, X. 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), Hong Kong, China, 3–7 November 2019; pp. 1718–1728. [\[CrossRef\]](#)
12. Mathur, P.; Sawhney, R.; Chopra, S.; Leekha, M.; Ratn Shah, R. Utilizing Temporal Psycholinguistic Cues for Suicidal Intent Estimation. In Proceedings of the Advances in Information Retrieval, Lisbon, Portugal, 14–17 April 2020; Jose, J.M., Yilmaz, E., Magalhães, J., Castells, P., Ferro, N., Silva, M.J., Martins, F., Eds.; Springer: Cham, Switzerland, 2020; pp. 265–271. [\[CrossRef\]](#)
13. Sawhney, R.; Joshi, H.; Flek, L.; Shah, R.R. PHASE: Learning Emotional Phase-aware Representations for Suicide Ideation Detection on Social Media. In Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, Virtual, 19–23 April 2021; pp. 2415–2428. [\[CrossRef\]](#)
14. Sawhney, R.; Joshi, H.; Gandhi, S.; Jin, D.; Shah, R.R. Robust suicide risk assessment on social media via deep adversarial learning. *J. Am. Med. Inform. Assoc.* **2021**, *28*, 1497–1506. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Sawhney, R.; Joshi, H.; Gandhi, S.; Shah, R.R. 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), Virtual, 16–20 November 2020; pp. 7685–7697. [\[CrossRef\]](#)
16. Shing, H.C.; Nair, S.; Zirikly, A.; Friedenber, M.; Daumé, H., III; Resnik, P. Expert, Crowdsourced, and Machine Assessment of Suicide Risk via Online Postings. In Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic, New Orleans, LA, USA, 5 June 2018; pp. 25–36. [\[CrossRef\]](#)
17. Sinha, P.P.; Mishra, R.; Sawhney, R.; Mahata, D.; Shah, R.R.; Liu, H. #suicidal—A Multipronged Approach to Identify and Explore Suicidal Ideation in Twitter. In Proceedings of the 28th ACM International Conference on Information and Knowledge Management, CIKM '19, New York, NY, USA, 3–7 November 2019; pp. 941–950. [\[CrossRef\]](#)
18. Zogan, H.; Razzak, I.; Jameel, S.; Xu, G. Hierarchical Convolutional Attention Network for Depression Detection on Social Media and Its Impact during Pandemic. *IEEE J. Biomed. Health Inform.* **2023**, *1*–9. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Sawhney, R.; Joshi, H.; Gandhi, S.; Shah, R.R. Towards Ordinal Suicide Ideation Detection on Social Media. In Proceedings of the 14th ACM International Conference on Web Search and Data Mining, WSDM '21, New York, NY, USA, 8–12 March 2021; pp. 22–30. [\[CrossRef\]](#)
20. Mishra, R.; Prakhar Sinha, P.; Sawhney, R.; Mahata, D.; Mathur, P.; Ratn Shah, R. SNAP-BATNET: Cascading Author Profiling and Social Network Graphs for Suicide Ideation Detection on Social Media. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Student Research Workshop, Minneapolis, MN, USA, 2–7 June 2019; pp. 147–156. [\[CrossRef\]](#)
21. Aragón, M.E.; López-Monroy, A.P.; González-Gurrola, L.C.; Montes-y Gómez, M. Detecting Depression in Social Media using Fine-Grained Emotions. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, MN, USA, 2–7 June 2019; Association for Computational Linguistics: Minneapolis, MN, USA, 2019; Volume 1 (Long and Short Papers), pp. 1481–1486. [\[CrossRef\]](#)
22. Gaur, M.; Alambo, A.; Sain, J.P.; Kursuncu, U.; Thirunarayan, K.; Kavuluru, R.; Sheth, A.; Welton, R.; Pathak, J. Knowledge-Aware Assessment of Severity of Suicide Risk for Early Intervention. In Proceedings of the World Wide Web Conference, WWW '19, New York, NY, USA, 13–17 May 2019; pp. 514–525. [\[CrossRef\]](#)
23. Mowery, D.L.; Park, A.; Bryan, C.; Conway, M. Towards Automatically Classifying Depressive Symptoms from Twitter Data for Population Health. In Proceedings of the Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media (PEOPLES), Osaka, Japan, 12 December 2016; pp. 182–191.
24. Uban, A.S.; Chulvi, B.; Rosso, P. An emotion and cognitive based analysis of mental health disorders from social media data. *Future Gener. Comput. Syst.* **2021**, *124*, 480–494. [\[CrossRef\]](#)

25. Zogan, H.; Razzak, I.; Wang, X.; Jameel, S.; Xu, G. Explainable depression detection with multi-aspect features using a hybrid deep learning model on social media. *World Wide Web* **2022**, *25*, 281–304. [[CrossRef](#)] [[PubMed](#)]
26. Ren, L.; Lin, H.; Xu, B.; Zhang, S.; Yang, L.; Sun, S. Depression Detection on Reddit with an Emotion-Based Attention Network: Algorithm Development and Validation. *JMIR Med Inf.* **2021**, *9*, e28754. [[CrossRef](#)]
27. Abdul-Mageed, M.; Ungar, L. EmoNet: Fine-Grained Emotion Detection with Gated Recurrent Neural Networks. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, Vancouver, BC, Canada, 30 July–4 August 2017; Barzilay, R., Kan, M.Y., Eds.; Association for Computational Linguistics: Minneapolis, MN, USA, 2019; Volume 1: Long Papers, pp. 718–728. [[CrossRef](#)]
28. Linmei, H.; Yang, T.; Shi, C.; Ji, H.; Li, X. Heterogeneous Graph Attention Networks for Semi-supervised Short Text Classification. 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), Hong Kong, China, 3–7 November 2019; pp. 4821–4830. [[CrossRef](#)]
29. Liu, T.; Hu, Y.; Wang, B.; Sun, Y.; Gao, J.; Yin, B. Hierarchical Graph Convolutional Networks for Structured Long Document Classification. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *34*, 1–15. [[CrossRef](#)]
30. Liang, B.; Su, H.; Gui, L.; Cambria, E.; Xu, R. Aspect-based sentiment analysis via affective knowledge enhanced graph convolutional networks. *Knowl.-Based Syst.* **2022**, *235*, 107643. [[CrossRef](#)]
31. Tang, H.; Ji, D.; Li, C.; Zhou, Q. Dependency Graph Enhanced Dual-transformer Structure for Aspect-based Sentiment Classification. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Virtual, 5–10 July 2020; pp. 6578–6588. [[CrossRef](#)]
32. Zhang, C.; Li, Q.; Song, D. Aspect-based Sentiment Classification with Aspect-specific Graph Convolutional Networks. *arXiv* **2019**, arXiv:cs.CL/1909.03477. [[CrossRef](#)]
33. Zhang, Y.; Yu, X.; Cui, Z.; Wu, S.; Wen, Z.; Wang, L. Every Document Owns Its Structure: Inductive Text Classification via Graph Neural Networks. *arXiv* **2020**, arXiv:2004.13826. [[CrossRef](#)]
34. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; Minneapolis, MN, USA, 2–7 June 2019; Burstein, J., Doran, C., Solorio, T., Eds.; Volume 1 (Long and Short Papers), pp. 4171–4186. [[CrossRef](#)]
35. Pennington, J.; Socher, R.; Manning, C. GloVe: Global Vectors for Word Representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; pp. 1532–1543. [[CrossRef](#)]
36. Han, S.C.; Yuan, Z.; Wang, K.; Long, S.; Poon, J. Understanding Graph Convolutional Networks for Text Classification. *arXiv* **2022**, arXiv:2203.16060.
37. Church, K.W.; Hanks, P. Word Association Norms, Mutual Information, and Lexicography. *Comput. Linguist.* **1990**, *16*, 22–29.
38. Long, S.; Cabral, R.; Poon, J.; Han, S.C. A Quantitative and Qualitative Analysis of Suicide Ideation Detection using Deep Learning. *arXiv* **2022**, arXiv:2206.08673. [[CrossRef](#)]
39. O’Dea, B.; Wan, S.; Batterham, P.J.; Calear, A.L.; Paris, C.; Christensen, H. Detecting suicidality on Twitter. *Internet Interv.* **2015**, *2*, 183–188. [[CrossRef](#)]
40. Posner, K.; Brent, D.; Lucas, C.; Gould, M.; Stanley, B.; Brown, G.; Fisher, P.; Zelazny, J.; Burke, A.; Oquendo, M.; et al. *Columbia-Suicide Severity Rating Scale (C-SSRS)*; Columbia University Medical Center: New York, NY, USA, 2008; Volume 10, p. 2008.
41. MacAvaney, S.; Mittu, A.; Coppersmith, G.; Leintz, J.; Resnik, P. Community-level Research on Suicidality Prediction in a Secure Environment: Overview of the CLPsych 2021 Shared Task. In Proceedings of the Seventh Workshop on Computational Linguistics and Clinical Psychology: Improving Access, Virtual, 11 June 2021; pp. 70–80. [[CrossRef](#)]
42. Mohammad, S.M.; Turney, P.D. Crowdsourcing a word–emotion association lexicon. *Comput. Intell.* **2013**, *29*, 436–465. [[CrossRef](#)]
43. Mohammad, S.M.; Kiritchenko, S. Using Hashtags to Capture Fine Emotion Categories from Tweets. *Comput. Intell.* **2015**, *31*, 301–326.
44. Cambria, E.; Liu, Q.; Decherchi, S.; Xing, F.; Kwok, K. SenticNet 7: A Commonsense-based Neurosymbolic AI Framework for Explainable Sentiment Analysis. In Proceedings of the Thirteenth Language Resources and Evaluation Conference, Marseille, France, 20–25 June 2022; Calzolari, N., Béchet, F., Blache, P., Choukri, K., Cieri, C., Declerck, T., Goggi, S., Isahara, H., Maegaard, B., Mariani, J., et al., Eds.; European Language Resources Association: Luxembourg, 2022; pp. 3829–3839.
45. Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; Stoyanov, V. Roberta: A robustly optimized bert pretraining approach. *arXiv* **2019**, arXiv:1907.11692. [[CrossRef](#)]
46. Ji, S.; Zhang, T.; Ansari, L.; Fu, J.; Tiwari, P.; Cambria, E. MentalBERT: Publicly Available Pretrained Language Models for Mental Healthcare. In Proceedings of the Thirteenth Language Resources and Evaluation Conference, Marseille, France, 20–25 June 2022; Calzolari, N., Béchet, F., Blache, P., Choukri, K., Cieri, C., Declerck, T., Goggi, S., Isahara, H., Maegaard, B., Mariani, J., et al., Eds.; European Language Resources Association: Luxembourg, 2022; pp. 7184–7190.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.