Hierarchical Convolutional Attention Network for Depression Detection on Social Media and Its Impact During Pandemic

Hamad Zogan, Imran Razzak, Shoaib Jameel, and Guandong Xu

***Abstract*— People across the globe have felt and are still going through the impact of COVID-19. Some of them share their feelings and suffering online via different online social media networks such as Twitter. Due to strict restrictions to reduce the spread of the novel virus, many people are forced to stay at home, which significantly impacts people's mental health. It is mainly because the pandemic has directly affected the lives of the people who were not allowed to leave home due to strict government restrictions. Researchers must mine the related human-generated data and get insights from it to influence government policies and address people's needs. In this paper, we study social media data to understand how COVID-19 has impacted people's depression. We share a large-scale COVID-19 dataset that can be used to analyze depression. We also have modeled the tweets of depressed and non-depressed users before and after the start of the COVID-19 pandemic. To this end, we developed a new approach based on Hierarchical Convolutional Neural Network (HCN) that extracts fine-grained and relevant content on user historical posts. HCN considers the hierarchical structure of user tweets and contains an attention mechanism that can locate the crucial words and tweets in a user document while also considering the context. Our new approach is capable of detecting depressed users occurring within the COVID-19 time frame. Our results on benchmark datasets show that many non-depressed people became depressed during the COVID-19 pandemic.**

***Index Terms*— Hierarchical CNN, Depression Detection, Twitter,**

Fig. 1: A sample of depressed user tweets during the first months of COVID-19.

I. Introduction

The novel coronavirus has quickly spread throughout the globe, affecting 218 nations and territories as depicted in Figure 2. By 31 July 2021, there were more than 4.2 million deaths and more than 197 million confirmed cases of illness.[1]. This new pandemic has created severe health and socio-economic disruptions worldwide. The World Health Organization (WHO) claims that COVID-19 is a pandemic that is taking lives throughout the globe every day, and the severity of the outbreak rates strongly supports this claim. The World Health Organization reported a new coronavirus outbreak in Wuhan, China, on 31 December 2019 (WHO) [1]. COVID-19 is spread by droplets and fomites during close, unprotected contact between sick and affected people who are not properly treated[2].

This epidemic has influenced mental health and the physical dangers posed by the virus [2]. The pandemic has not only caused complications by physically affecting the individuals but also impacted their mental lives. The lockdown and stay-at-home policies have also affected different types of individuals; for instance, a recent study found that lockdown could lead to more violence in the UK[3], which will have a detrimental societal impact. After staying at home for very long, quickly lifting lockdown restrictions will have an impact on the mental health too[4]. According to some studies [3]–[6], the world's population will face many psychological problems related to anger, anxiety, and others due to their isolation in homes. These studies point out that people can only get out of this psychological state initially through the intervention of governments in psychological support during crises, not only treating them after disasters are over. With the spread of COVID-19, mental health problems are suspected of increasing among people. Other cascading factors that could severely impact people's mental health are job losses, increasing prices, non-availability of certain food items, and others.

Mental health distress, including peri-quarantine anxiety, depression, and phobias specific to COVID-19, is supposed to increase during the pandemic, affecting individuals and calling for intervention. Due to pandemic-induced social distancing protocols, health interventions have become the most common form of psychosocial support for mental health. However, there is no evidence of the ef-

Hamad Zogan is currently pursuing the Ph.D. degree with the University of Technology Sydney, Ultimo, NSW, Australia, with a full scholarship from Jazan University, Jizan, Saudi Arabia.
Imran Razzak is with School of Computer Science and Engineering, University of New South Wales, Sydney, Australia
Shoaib Jameel is with School of Electronics and Computer Science, University of Southampton, UK
Guandong is are with School of Computer Science and Data Science, University of Technology Sydney, Australia: Corresponding Author: Guandong Xu, guandong.xu@uts.edu.au

[1]https://covid19.who.int/

[2]https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf

[3]https://www.msn.com/en-gb/health/mindandbody/youth-violence-likely-to-explode-over-summer-uk-experts-fear/ar-AAMtHuO?ocid=msedgntp

[4]https://www.msn.com/en-gb/health/mindandbody/lifting-restrictions-quickly-is-just-as-detrimental-to-our-mental-health-as-lockdown/ar-AAMvF5f?ocid=msedgntp

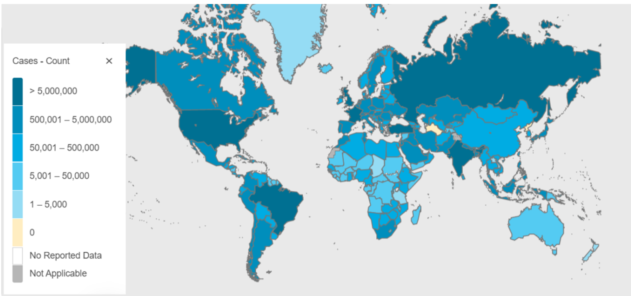


Fig. 2: As of July 23, 2021, the following countries or geographic locations have confirmed cases of COVID-19.

fectiveness of remote psychosocial interventions in improving mental health outcomes. Infection control strategies such as lockdown and social distancing (physical distancing) try to slow disease spread by decreasing direct contact among people. Conversely, lockdown, quarantine, and social isolation may lead to personal freedom loss, future uncertainty and anxiety, as well as a large increase in the incidence of emotional diseases and mental health problems [7], [8]. Lockdown measures are connected to higher than average incidence of mental illness such as depression, according to a study by Rossi et al., [9]. More extended quarantine periods, fears of illness, frustration, boredom, insufficient supplies, lack of knowledge, financial loss, and stigma were all stressors. As a result, Tull et al., [10] found out that remaining at home was linked to higher levels of health anxiety, financial anxiety, and loneliness among healthy people.

In the midst of the COVID-19 epidemic, social media usage has increased as more people rely on acquiring the newest concerning COVID-19 information [11]. Additionally, social media, including a variety of internet services and platforms like Facebook and Twitter, allow users to converse, interact and different information sharing. While social media can help distribute information, which could be beneficial in the fight against the epidemic, it has also been related to anxiety and depression [12]. We hypothesize that the COVID-19 pandemic and its social restrictions may affect depressed users and thus may be reflected in their daily tweets. Figure 1 shows a sample of tweets from a depressed user during this pandemic. From the Figure, we notice that the user continues to post about their mental health issue, and we believe there is an increase in the proportion of depression-related tweets during the pandemic compared to before. In April 2020, several countries performed various lockdown restrictions and testing strategies to contain the virus's spread [13], and these restrictions were reflected in user tweets, as we observe in Figure 1. One of the recent studies by Hua et al., [14], shows that many people use social media to disseminate terrifying COVID-19-related information, such as information on individuals who passed away as a result of COVID-19 or who are battling the disease; hence, many of these posts could be shared via depressed people.

In January 2020, lockdowns for the COVID-19 pandemic began in China [13]. They were quickly followed by lockdowns in many countries worldwide, including Egypt and Germany. The first lockdown, crucially, was generally unanticipated by the general public in all countries. As a result, not everyone can tolerate the limits in daily life and behavior equally well. WHO and health care professionals have recommended that counseling programs supporting and directing individuals in behavior modification become part of the COVID-19 pandemic preventive measures to avoid extra mental health costs in the general population [15]. However, to properly promote mental health, well-being, and behavior, a better scientific knowledge of how individuals perceive and mentally react to the present COVID-19 epidemic is needed. A better scientific understanding of the social aspects of life most affected by the current COVID-19 pandemic, including how individuals think, feel, suffer, deal with situations, and perceive a threat, emotion, and emotion behavior control [16]. Moreover, a depressed user may post in various linguistic styles, using depressing words, antidepressant mentions, and descriptions of depressed symptoms, all of which can help diagnose depression. While tweets may contain rich information sources, and due to their unstructured nature, obtaining insights from them can be difficult and time-consuming. Organizing text data is becoming ideal due to advancements in natural language processing (NLP) [17]. In particular, text classification is among the most fundamental techniques in machine learning. In text classification, given annotated training data, the model learns patterns from this data. Given some unseen data, the goal of the model is to assign labels to each instance; annotations or labels could be in the form of topic, genre, or sentiment [18]–[20]. The importance of text classification has never been greater than it is now, especially for companies seeking to boost productivity or profitability. Most recently, new deep learning models have achieved SOTA performance in text classification. More recently, transfer learning methods have become very popular. In this setup, a model is first trained on some general text collection. The model is then fine-tuned on task-specific problems in certain domains. Several models already have been developed that belong to such as class of transfer learning model, for instance, ULMFit fine-tunes a 3-layered LSTM [21] and Bidirectional Encoder Representations from Transformers (BERT) from Google [22] has been shown to be effective in several downstream tasks. Nevertheless, classifying tweets in social media is still challenging due to the structure of the user tweets and the fact that these text instances are very short and sometimes ambiguous.

Recently, the hierarchical attention networks (HAN) model proposed by Yang et al., [23] achieved SOTA results on text classification. What differentiates HAN from other existing approaches to document classification is that it exploits the rchical nature of text data by dividing a document into two levels, word-level hierarchy and sentence-level hierarchy. The authors proposed a new model that maintains a hierarchical structure consisting of word and sentence encoders. The word encoder extract features at the word level and sends them to the sentence encoder for processing. The sentence encoder extracts the features at the sentence level and predicts output probabilities at the final layer. However, HAN takes substantially longer to train than CNN-based techniques because they utilize RNNs. Long-distance linguistic elements are more difficult to capture using CNNs. Despite this shortcoming, CNNs are typically just as efficient as RNNs in performing many basic NLP tasks. The reason why we developed our novel model based on HAN is that user posts may include linguistic cues at various levels of word and tweet levels. To identify a depressed user on social media, every word in a tweet and every user's tweet is equally essential Figure 1.

We propose a new hierarchical convolutional neural network (HCN) model to better model user tweet classification through social media. Our model can extract meaningful feature representations from the word and tweet levels. Eventually, we studied our model performance using a multi-channel convolution neural network CNN and multilayer perceptron MLP in word-level encoding to combine the advantages of two traditional neural network models. We found our model performance increased when we used two channels for word encoding, and we call this model with two channels HCN+. Our research makes the following keys contributions:

- We develop a depressed user classification model based on a hierarchical convolution neural network **(HCN)** and hierarchical attention that can learn and extract linguistic relationships and local features from tweets.

- Besides HCN, we present a variant of hierarchical convolution neural network **HCN+**. The model (HCN+) expands word encoding by using two channels and reading the input differently in parallel using multilayer perceptron and convolutional neural network to extract various features from the same input and boost our model performance to identify depressed Twitter users.
- We collected a dataset, including posts from users with and without depression during COVID-19, to identify and analyze online depression during the pandemic. We are aware that this is the first study evaluating the behavior of depressed users during the time of the COVID-19 epidemic. Furthermore, we have made this dataset publicly available to aid in the research on mental health and well-being[5].

II. Related Work

In this section, we summarise some closely related work, including our previous works. We mention how the model developed in this work is substantially novel and different from our previous works.

A. Depression Detection on Social Media

The majority of prior research, including ours, have focused on user behaviour to identify whether the user is depressed or suffering from another mental disease. Deep learning approaches for online depression detection, such as public discussion forums, have been developed to model data from the social media. Most methods take words and word n-grams as user-level features and apply classic classification algorithms such as Support Vector Machines (SVM).

Users that were depressed had much greater ratios of negative emotion phrases, according to Tsugawa et al [24]. In general, the linguistic style of postings may be used to extract features [25], [26], [26]. Many factors, such as user tweets, replies, retweets, post time, emotions, and so on, might help identify depression. Shen et al. [27] used Twitter to create a well-labelled depression and non-depression dataset and then investigated a variety of depression-related features and developed a multimodal depressive model to identify depressed users. They grouped these features into six depression-related features, including clinical depression criteria and online social media behaviours.

Wang et al., [28] proposed a way for automatically gathering people who described themselves as having an eating disorder in their Twitter profile descriptions to detect eating disorders within social media communities. The authors collected features of linguistic from users for psychometric qualities, and they used similar settings described in [29]–[31]. From Twitter and Weibo, the authors collected 70 features. They took these characteristics from a user's profile, and user engagement characteristics like many followees and followers. Wong et al., [28], on the other hand, integrated user-level and post-level semantics and framed their problem as many instances learning setup; this approach has the benefit of learning from user-level labels to post-level labels identification [32]. Moreover, due to the need to describe the underlying workings of a deep learning system and make them more dependable and understandable, recently explainable machine learning has drawn a lot of attention [23], [33], [34]. Inspired by that, Zogan et al., [35] recently introduced a model that improves depression detection by incorporating the explainability of depressed user tweets. Their model interplay between multilayer perceptron (MLP) and hierarchical attention network (HAN). MLP was used to encode users' online behaviour, while HAN encoded all user tweets at two levels: word-level and tweet-level. They determined each tweet and word weight and extracted characteristics derived from user tweets' semantic sequences. Another work by [36], the authors argued that using all user tweets to identify a depressed user is ineffective and could even degrade a model performance; therefore, they proposed a new summarization framework interplay between extractive and abstractive summarization to generate a shorter representation of user historical tweets and help to reduce the influence of content that may not eventually benefit the classifier.

B. Depression Detection due to COVID-19

COVID-19 has had a negative influence on people's mental health, and as a result of the epidemic, many people are experiencing an increase in mental health issues. Various studies have recently reported this. For instance, Zhou et al., [37] developed a depression classification model that extracts multimodal features from topic, emotion and domain-specific viewpoints using the tf-idf algorithm. Their study focuses on detecting community level depression in Australia and local areas only during the pandemic. Their findings reveal an increase in depression following the COVID-19 epidemic. While in our new study, we will analyse tweets of users with and without depression during eight months before and after the start of the COVID-19 pandemic. Galea et al., [38] claimed that the pandemic's mental health repercussions are applicable in the short-term and long-term, and that COVID-19 provided an opportunity to explore overseas students' behavioural responses to the imposed limits on travel and social interaction. Due to social distancing, the current COVID-19 epidemic substantially impacts daily activities, including work, social, educational, and recreational activities. Some of these situations may raise the risk of aggressiveness and suicide [39], [40]. During the pandemic period in China, Huang et al., [39] investigated anxiety disorder, depression, and sleep quality and found that the participants had varying degrees of mental health difficulties. According to the same study results, 35% of participants had anxiety, 20% had depression, and 18% had poor sleep quality.

Our proposed approach, on the other hand, focuses on automatically identifying depression online, which has the potential to monitor people's mental health. We have also developed a novel model for automatic modelling on user's depression as a result of the pandemic.

III. Dataset

To study depressed users during the COVID-19 pandemic, we have created a new dataset that includs users tweets pre- and during COVID-19. We show in Table I a summary of this dataset.

	Pre-COVID-19		**COVID-19**	
	# User	# Tweet	# User	# Tweet
Depressed	2,159	447,856	741	326,129
Non-Depressed	2,049	1,349447	682	931,527

TABLE I: Summary of the datasets that we used in our research

A. Pre-Covid-19 Dataset

Shen et al., [27] constructed a textual depression dataset on Twitter. The authors labeled users as depressed if they found a user's tweet that contained a specific pattern. Between 2009 and 2016, they constructed depressed users based on the content of their posts. There are around 2K depressed users and around 400K tweets in all. The tweets for non-depressed were collected in December 2016, including over 2K users and over a million tweets.

[5]https://github.com/hzogan/Post-COVID_dataset

B. COVID-19 Dataset

Here, we provide a description of our dataset, which was generated following the COVID-19 pandemic started to penetrate across different nations. according to the tweets' IDs in [27], first, we constructed user tweets after COVID-19 using Twitter APIs. We set up a time-frame to study all users in the COVID dataset from the 1st of September 2019 until the 20th of April 2020. We crawled tweets of 1423 users who were tweeting during the time-frame of our study. Finally, we constructed this novel dataset with over a million and two hundred tweets from both depressed and non-depressed users based on these tweets.

IV. Our Proposed Model

Figure 3 depicts the hierarchical attention network that was provided to learn user tweets representation of as we were inspired by [23]. As mentioned above, there are some key differences between our model and the HAN model. One difference is that we have used the CNN model in our framework with two channels for word encoding. Assume that U is a user who posted M tweets $T = [t_1, t_2, ..., t_M]$ each of which had N_i words $t_i = [w_1, w_2, ..., w_N]$. The series of d-dimensional embeddings of each tweet's words, $w_i = [w_{11}, ..., w_{MN}]$, serves as its representation. Each word is represented by a fixed-size vector from pre-trained word embeddings as the input layer. Figure 3 depicts the overall structure of our HCN+.

A. Word Encoding

Given a tweet with words w_{it}, $t \in [0, T]$, we begin by converting the words to vectors using an embedding matrix. Each word vector is represented by a fixed-size vector obtained using pre-trained word embeddings as the input layer. To initially capture the contextual information of the annotations, CNN is employed as the word level encoder. We utilized one-dimensional convolution, where a filter vector slides across a sequence and detects features at different points in one-dimensional convolution. The convolution kernel completes the convolution operation in the k-dimensional window, the convolution kernel $w \in \mathbb{R}^{hk}$ operates on the input data matrix $x_{i:i+h-1}$ to create the feature c_i each time, and the output features are as follows:

$$\mathrm{c}_i = f(wx_{i:i+h-1} + b) \tag{2}$$

where f is the nonlinear activation function $relu$, h is the sliding window range, b represent the bias and $x_{i:i+h-1}$ is the result of joining the word x_i to x_{i+h-1}. As a result, the convolution kernel is applied to the tweet sequence of length n, yielding $n-h+1$ results:

$$c = \{c_1,\ c_2,\ \cdots,\ c_{n-h+1}\} \tag{10}$$

Reducing the size of features is done using the pooling layer and eliminate redundant information. The pooling process is carried out using the pooling technique. The K-max pooling layer is utilised in our architecture to down-sample the local feature maps C and extract global feature representations of short texts with fixed-length, as follows:

$$\hat{c} = \max(c) \tag{3}$$

The attention mechanism is then described. It is important to create a trainable and expected to capture global word vector u_{ij} for all words. The $\hat{c}_{ij}$ annotations build the basis for attention, which begins with another hidden layer. The annotations u_{ij} will be represented as following:

$$u_{ij} = \tanh(W_w \hat{c}_{ij} + b_w) \tag{1}$$

Where W_w and b_w are the trainable parameters that learned by the model after random initialisation. Then the product $u_{ij}u_w$ (u_w is randomly initialised) is intended to emphasise the significance of the word j and normalised by a softmax function to an importance weight per word α_{ij}:

$$\alpha_{ij} = \frac{\exp(u_{ij}u_w)}{\sum_j \exp(u_{ij}u_w)} \tag{2}$$

A weighted concatenation of word representations and the previously determined annotations is known as the tweet vector v_i, where α_t is the importance weight per word:

$$v_i = \sum\nolimits_t \alpha_{ij}\hat{c}_{ij} \tag{3}$$

For our first model HCN, we have utilized one channel CNN for word encoding Figure 4 depicts the word encoding component for HCN. Then, in order to combine the advantages of two traditional neural network models and alleviate their shortcomings, we have utilized two channels CNN and a multilayer perceptron (MLP), to obtain multi-channel representations. The multi-channel representations reflect the various contributions of different words to a tweet's semantics and provide a way to represent a tweet from several views. Hence, we have examined the effectiveness of using multi-channel CNN-MLP for the word encoding of our second model HCN+ depicted in Figure 5. Each deep learning model has its own method for converting target data into feature vectors. The embedding layer first fed into the MLP with the attention model, which generates an MLP feature vector and fed CNN with the attention model, generating a CNN feature vector. These vectors are concatenated and produced a tweet embedding representing the content within that tweet.

B. Tweet Encoder

Yang et al. [23], achieved SOTA performance on HANs by employing a hierarchical framework that breaks down texts into sentences first. The word-level reads in word embeddings from a given sentence. It produces a sentence embedding representing the content within that sentence. In comparison, the tweet-level reads in the tweet embeddings produced by the word-level as depicted in Figure 6. Given that a user's tweets might not be significant in identifying and elucidating a depression of a person, the attention layer is employed with the tweets of user to obtain the more important tweets. Utilizing the tweet level attention layer, we will gather the connected tweets. The importance of the i tweet is predicted to be indicated by the product $u_i u_s$, which has been standardised to a α_i importance weight for each tweet. Eventually, a vector called $\hat{t}$ will be used to compile all the tweet data from user posts:

$$\hat{t} = \sum\nolimits_t \alpha_i h_i^t \tag{4}$$

C. Classification Layer

At the classification layer, we must determine whether or not the user is depressed. So far,we've covered the encoding of user tweets (s) by simulating the tweet and word levels of the hierarchical structure. Such a network's output is often sent to a sigmoid layer for classification:

$$\hat{y} = \mathrm{Sigmoid}(b_f + \hat{t}W_f) \tag{5}$$

where $\hat{y}$ denotes the predicted probability vector. Predicted probabilities of labels being 0 (non-depressed user) and 1 (depressed user), they are correspondingly represented by the symbols $\hat{y}_0$ and

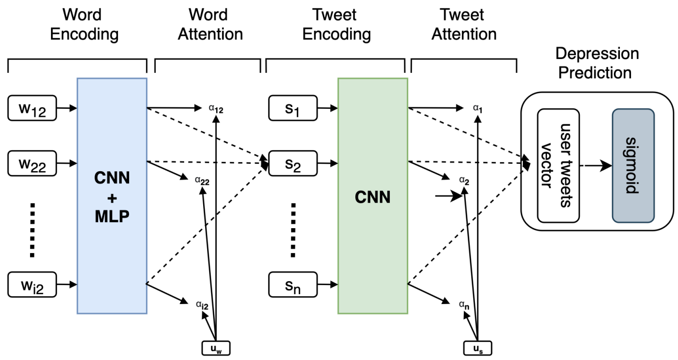


Fig. 3: A diagram of (HCN) that we employed for user all user posts

TABLE II: Performance Comparison on Pre-COVID datasets. HCN+ outperforms baselines.

Training Data	Model	Precision	Recall	F1-Score	Accuracy
Tweets Summarization	XLNet (base)	0.889	0.808	0.847	0.847
	BERT (base)	0.903	0.77	0.831	0.837
	RoBERTa (base)	**0.941**	0.731	0.823	0.836
	BiGRU-Att	0.861	0.843	0.835	0.837
	CNN-Att	0.836	0.829	0.824	0.824
	CNN_BiGRU-Att	0.868	0.843	0.848	0.835
All user tweets	BiGRU	0.766	0.762	0.786	0.764
	CNN	0.817	0.804	0.786	0.806
	HAN	0.870	0.844	0.856	0.835
	HCN	0.853	0.852	0.852	0.852
	HCN+	0.871	**0.868**	**0.869**	**0.869**

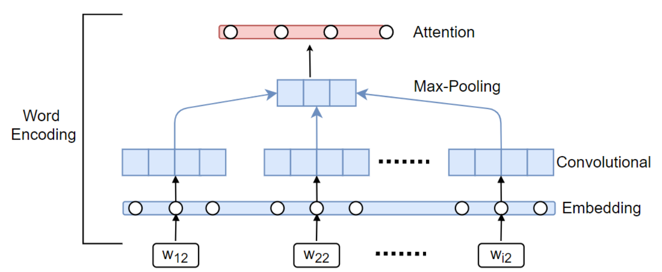


Fig. 4: An illustration of one channel CNN model that we use for HCN word encoding

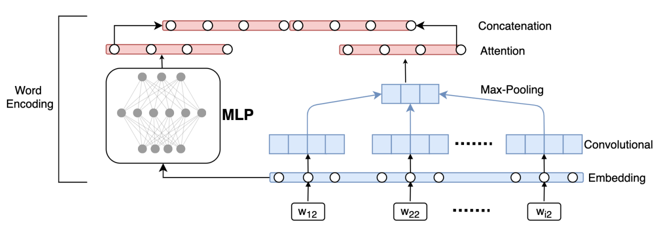


Fig. 5: An illustration of two channel CNN+MLP model that use for HCN+ word encoding

$\hat{y_1}$. Afterward, the ground-truth label for each user y, we seek to minimise the cross-entropy error:

$$\text{Loss} = -\sum_i y_i \cdot \log \hat{y}_i \tag{6}$$

y_i represents the user who has the ground truth label (either non-depression or depression) and $\hat{y}_i$ is the predicted probability.

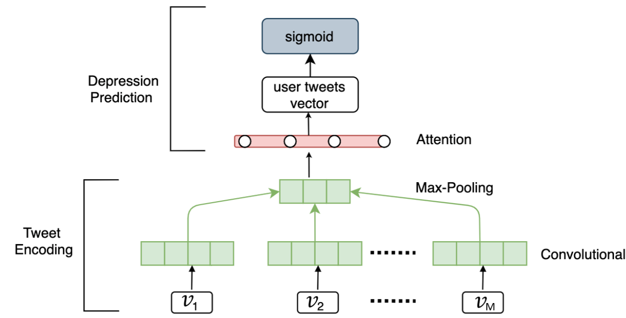


Fig. 6: An illustration of a tweet encoder network

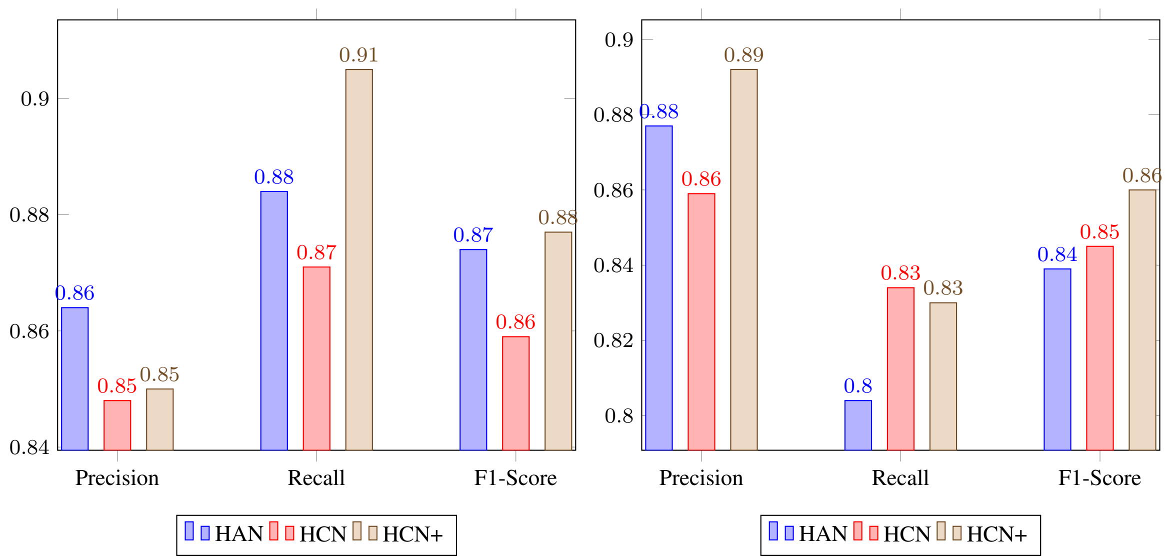
V. Experiments and Results

The experimental results of our model are shown in this section and compares them with different comparative models. We also present qualitative results.

A. Experiment Setup

1) Dataset: To evaluate the effectiveness our models, we conduct our experiments on Shen et al., [27] pre-COVID dataset, as shown in Table I, which contains users and their posts on Twitter. Each user is labelled either depressed or non-depressed. For preprocessing, we remove users who have less than ten tweets and for evaluation, we randomly split the dataset into training and test set with a ratio of 80:20 with 5-fold cross-validation.

2) Comparative methods: This section presents an experimental assessment to verify HCN's performance. We looked into three well-known pre-trained models since they are frequently used in contextual language models created using recent deep learning techniques. Our model is compared with different classification methods using the following data inputs:



(a) Depression (b) Non-depression

Fig. 7: Comparison between HCN+ and other hierarchical text classification models (a) for depression prediction and (b) for nod-depression prediction

1) **Tweets Summarization**: Zogan et al, [36] summarize all user tweets utilizing a new summarization framework interplay between extractive and abstractive summarization to generate a shorter representation of user historical tweets and help to reduce the influence of content. Their experiments for summarization sequence classification have examined several models, Conventional neural network with attention CNN-Att, Bidirectional Gated Recurrent Neural Network with Attention BiGRU-Att. Three pre-trained models for transformers have also been investigated by the authors, and they are XLNet [41], BERT [22] and RoBERTa [42].
2) **All User Tweets**: For all user tweets, our model is comapred to the following classification methods:
   - **BiGRU**:For the purpose of classifying user tweets, we employed the **Bidirectional-GRU** [43] with attention method that we deployed to get user tweet representations.
   - **CNN**: In order to represent user tweets and capture the semantics of various convolutional window sizes for depression detection, we used **CNN** [44] with an attention mechanism.
   - **HAN**:In order to identify depression in user postings, a hierarchical attention neural network architecture [23] is deployed. The network encrypts first user postings by paying attention to both the words in each tweet and the tweets themselves. Bidirectional-GRU is used (GRU).
   - **HCN**: Similar to HAN, however instead of utilizing (GRU), hierarchical convolutional network (HCN) rely on architectures based off convolutional neural networks.
   - **HCN+**: The proposed model in this paper.

3) Evaluation metrics: We employ the standard metrics for information retrieval such as accuracy, F1-score, and precision as metrics to examine the classification performance. These metrics are widely used in previous works for depression detection [27], [35]–[37].

4) Experimental settings: We performed the experiments with Python 3.6.3 and Tensorflow 2.1.0. Word embeddings is initialized by Glove [45]. The dimension of word embedding is 100, and the dropout rate is set to 0.5. We created 100 different filters for the convolutions, each with length 4, so the result will bring 100 different convolutions. We used the Adam optimization algorithm for both HCN and HCN+ with default value learning rate (lr) = 0.001. We train HCN+ for 20 epochs on all the data with a batch size of 32.

B. Results

In this section, we report the quantitative results obtained from different models. Evaluation results for different competing methods are presented, where the best results for the best model are highlighted in bold in Table II. The first part of the table shows the effective results of using summarization sequences of user posts to detect depression; the performance compared using some models that achieved a new state-of-the-art result on many NLP tasks, such as text classification. We see that CNN_BiGRU-At outperforms other models with F1-score and recall; however, XLNet and RoBERTa perform best among all the different models with accuracy and precision, respectively.

The second part of Table II, shows the results of using all user tweets; we observed that all the hierarchical text classification models (HAN, HCN and HCN+) efficient outperform other neural network-based methods, such as BiGRU and CNN. Furthermore, we observed that our two models, HCN and HCN+, outperform other models in terms of Accuracy and Recall. Comparing our model with HAN, Our model HCN+ boosts about 3.4%, 1.3%, 2.4% and 0.1% in terms of Accuracy, F1-score, Recall and Precision.

Generally, our models based on the hierarchical network can consistently outperform other methods in terms of Accuracy, F1 Score and Recall on both training data (tweets summarization and all user tweets). Our models based on hierarchical networks achieve a relative improvement of 0.5% for HCN and 2.2% for HCN+, compared against the best results (XLNet) in terms of Accuracy.

Finally, to better understand and examine the effectiveness of our model, we have compared the performance of our model (HCN+) with other hierarchical text classification models (HAN and HCN) in order to predict the depressed (Figure 7-a) and non-depressed-users (Figure 7-b). We found out HCN+ obtains better performance in both labels in terms of F1-score, showing 87% and 86% for depressed and non-depressed users, respectively.

C. Discussion

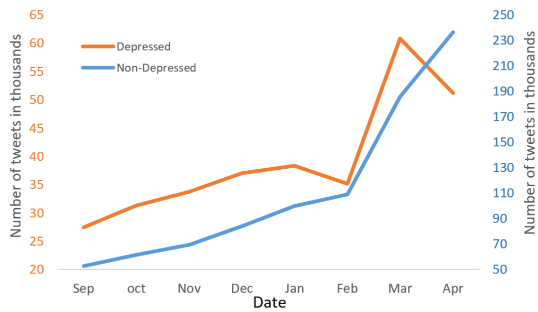


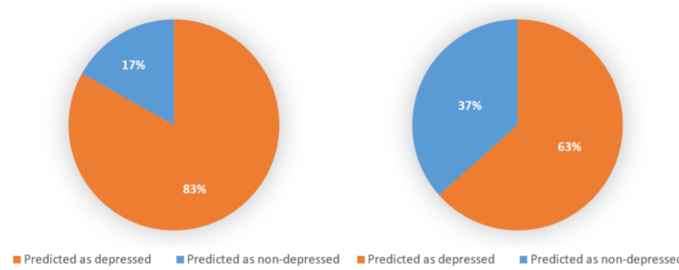
Fig. 8: Monthly tweets for all users during the COVID-19

To study user tweets dynamic during COVID-19, we set up a time window to study all users in the pre-COVID dataset from 1 September 2019 until 20 April 2020. According to the table, we studied 1423 users; 741 were labeled as depressed users before COVID-19. Figure 8 shows the monthly users' posts during our study window. We see that tweets for both depressed and non-depressed users have increased per month since the first cases of COVID-19 were reported (according to Lancet journal [1]). We have also observed that in February 2020, the number of tweets increased dramatically among all users.

After having testified our classification model, we utilize HCN+ to users in the COVID dataset (Table III). First, we studied all user-collected tweets in the time frame from September to April and showed the proportion of positive cases among depressed and non-depressed users during COVID-19 Figure 9. As shown in Figure 9a, more than 80% of depressed users were predicted via HCN+ as depressed during the four-months time frame. The proportion in Figure 9 looks natural since these users have been diagnosed with depression before COVID-19. However, Figure 9b shows that 63% of non-depressed users were predicted as depressed. It shows how COVID-19 impacted normal users' tweets during the pandemic.

We further analyzed monthly dynamic user tweets. Figure 10 shows monthly dynamic depressed user tweets during our study time frame. We observed the number of depressed users predicted as positive, roughly the same during the first three months. However, once the first case was reported in late December 2019 in Wuhan [1], we can see that the number of positive cases decreased in December, which may be due to depressed users posting news tweets instead of tweets to expose their feelings. We also noticed that the number of positive cases of depressed users reached a peak in March, which is the same month that World Health Organization WHO declared that COVID-19 is a pandemic [46]. Also, in March, several countries performed various lockdown restrictions and testing strategies to contain the virus's spread [13].

On the other hand, non-depressed dynamic tweets in the first four months of the study time-frame are quite similar to depressed users' tweets during the same time frame Figure 10. Like depressed users, the number of positive cases of non-depressed users also peaked in March. However, we noticed something considerably interesting, the portion pattern of depressed users' positive cases before and after December was slightly the same. While in non-depressed users, as we can see in Figure 11, the number of positive cases rate starting to increase in January compared to the positive cases rate in November, which is the month before the first case of Corona was announced. We found out the positive cases rate increased by 3%, 9%, 15% and 12% in January, February, March and April, respectively.



(a) Depressed Users (b) Non-depressed Users

Fig. 9: The proportion of positive cases among depressed and non-depressed users during COVID-19.

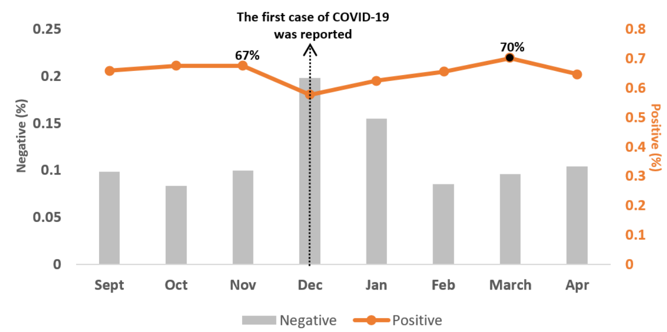


Fig. 10: Depressed user dynamics between September 1, 2019, and April 20, 2020

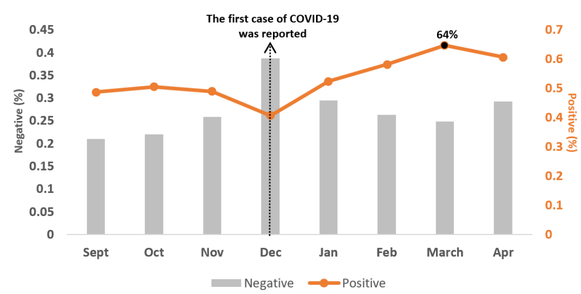


Fig. 11: Non-depressed user dynamics between September 1, 2019, and April 20, 2020

VI. Conclusion

In this paper, we studied tweets of depressed and non-depressed users during eight months before and after the start of the COVID-19 pandemic. A user classification model to automatically detect depressed users based on a hierarchical convolution neural network (HCN) is proposed, which exploits data from Twitter. HCN considers the hierarchical structure of user tweets (tweets-words) and contains an attention mechanism that can find the most crucial tweets and words in a user document while also considering the context. We expand word encoding by using two channels and read the input in different ways in parallel using MLP and CNN to extract different features from the same input and boost our model performance. The results showed that our two models (HCN and HCN+) outperform strong comparative models and effectively detect depressed users. Furthermore, we have also investigated depressed users' well-being during COVID-19 by crawling and classifying their tweets during the pandemic. Moreover, we have analyzed depressed users' well-being during COVID-19 by crawling and classifying their tweets during the pandemic. We found that the COVID-19 pandemic and its restrictions, such as lockdowns and changes in the workplace, impacted many depressed and non-depressed users. In the future, besides the users' tweets, we will analyze user behaviors related to depression during the COVID pandemic, such as social engagement and social interaction with others. This would provide the model with more contextual information and allow us to concentrate on a task where our model not only detects depression but also automatically gives a possible diagnosis. Moreover, we will aim to detect a user's loneliness during the pandemic, which is one mental illness that has never been before in this field. Loneliness is considered one of the early depression symptoms; therefore, its detection will help in early depression detection.

VII. Acknowledgement

This work is partially supported by the Australian Research Council (ARC) under Grant No. DP200101374 and LP170100891.

References

[1] Boaz Shmueli. Multi-class metrics made simple, part ii: the f1-score. *Retrieved from Towards Data Science: https://towardsdatascience.com/multi-class-metrics-made-simplepart-ii-the-f1-score-ebe8b2c2ca1*, 2019.
[2] World Health Organization". Mental health and covid-19.
[3] Kapil Goyal, Poonam Chauhan, Komal Chhikara, Parakriti Gupta, and Mini P Singh. Fear of covid 2019: First suicidal case in india! 2020.
[4] Lijun Kang, Simeng Ma, Min Chen, Jun Yang, Ying Wang, Ruiting Li, Lihua Yao, Hanping Bai, Zhongxiang Cai, Bing Xiang Yang, et al. Impact on mental health and perceptions of psychological care among medical and nursing staff in wuhan during the 2019 novel coronavirus disease outbreak: A cross-sectional study. *Brain, behavior, and immunity*, 87:11–17, 2020.
[5] Xin Shen, Xiaoyue Zou, Xiaofeng Zhong, Jing Yan, and Li Li. Psychological stress of icu nurses in the time of covid-19, 2020.
[6] Edmond Pui Hang Choi, Bryant Pui Hung Hui, and Eric Yuk Fai Wan. Depression and anxiety in hong kong during covid-19. *International journal of environmental research and public health*, 17(10):3740, 2020.
[7] Bazghina-werq Semo and Souci Mogga Frissa. The mental health impact of the covid-19 pandemic: implications for sub-saharan africa. *Psychology Research and Behavior Management*, 13:713, 2020.
[8] Bradford P Wilson. The phenomenon of grade inflation in higher education. In *Phi Kappa Phi Forum*, volume 79, page 38. National Forum: Phi Kappa Phi Journal, 1999.
[9] R Rossi, V Socci, D Talevi, S Mensi, C Niolu, F Pacitti, A Di Marco, A Rossi, A Siracusano, and G Di Lorenzo. Covid-19 pandemic and lockdown measures impact on mental health among the general population in italy. front psychiatry. 2020; 11: 790, 2020.
[10] Matthew T Tull, Keith A Edmonds, Kayla M Scamaldo, Julia R Richmond, Jason P Rose, and Kim L Gratz. Psychological outcomes associated with stay-at-home orders and the perceived impact of covid-19 on daily life. *Psychiatry research*, 289:113098, 2020.
[11] Gordon Pennycook, Jonathon McPhetres, Yunhao Zhang, Jackson G Lu, and David G Rand. Fighting covid-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention. *Psychological science*, 31(7):770–780, 2020.
[12] Edmond Ng. The pandemic of hate is giving covid-19 a helping hand. *The American journal of tropical medicine and hygiene*, 102(6):1158, 2020.
[13] David Koh. Covid-19 lockdowns throughout the world. *Occupational Medicine*, 70(5):322–322, 2020.
[14] Jinling Hua and Rajib Shaw. Corona virus (covid-19) "infodemic" and emerging issues through a data lens: The case of china. *International journal of environmental research and public health*, 17(7):2309, 2020.
[15] World Health Organization. Depression.
[16] Jay J Van Bavel, Katherine Baicker, Paulo S Boggio, Valerio Capraro, Aleksandra Cichocka, Mina Cikara, Molly J Crockett, Alia J Crum, Karen M Douglas, James N Druckman, et al. Using social and behavioural science to support covid-19 pandemic response. *Nature human behaviour*, 4(5):460–471, 2020.
[17] Christopher Manning and Hinrich Schutze. *Foundations of statistical natural language processing*. MIT press, 1999.
[18] Rushlene Kaur Bakshi, Navneet Kaur, Ravneet Kaur, and Gurpreet Kaur. Opinion mining and sentiment analysis. In *2016 3rd international conference on computing for sustainable global development (INDIACom)*, pages 452–455. IEEE, 2016.
[19] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. Thumbs up? sentiment classification using machine learning techniques. *arXiv preprint cs/0205070*, 2002.
[20] Peter D Turney. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. *arXiv preprint cs/0212032*, 2002.
[21] Jeremy Howard and Sebastian Ruder. Universal language model fine-tuning for text classification. *arXiv preprint arXiv:1801.06146*, 2018.
[22] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.
[23] Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. Hierarchical attention networks for document classification. In *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies*, pages 1480–1489, 2016.
[24] Sho Tsugawa, Yusuke Kikuchi, Fumio Kishino, Kosuke Nakajima, Yuichi Itoh, and Hiroyuki Ohsaki. Recognizing depression from twitter activity. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, pages 3187–3196. ACM, 2015.
[25] Quan Hu, Ang Li, Fei Heng, Jianpeng Li, and Tingshao Zhu. Predicting depression of social media user on different observation windows. In *2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, volume 1, pages 361–364. IEEE, 2015.
[26] Amir Hossein Yazdavar, Hussein S Al-Olimat, Monireh Ebrahimi, Goonmeet Bajaj, Tanvi Banerjee, Krishnaprasad Thirunarayan, Jyotishman Pathak, and Amit Sheth. Semi-supervised approach to monitoring clinical depressive symptoms in social media. In *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*, pages 1191–1198. ACM, 2017.
[27] Guangyao Shen, Jia Jia, Liqiang Nie, Fuli Feng, Cunjun Zhang, Tianrui Hu, Tat-Seng Chua, and Wenwu Zhu. Depression detection via harvesting social media: A multimodal dictionary learning solution. In *IJCAI*, pages 3838–3844, 2017.
[28] Tao Wang, Markus Brede, Antonella Ianni, and Emmanouil Mentzakis. Detecting and characterizing eating-disorder communities on social media. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, pages 91–100. ACM, 2017.
[29] Diana Ramírez-Cifuentes, Marc Mayans, and Ana Freire. Early risk detection of anorexia on social media. In *International Conference on Internet Science*, pages 3–14. Springer, 2018.
[30] Akshi Kumar, Aditi Sharma, and Anshika Arora. Anxious depression prediction in real-time social data. *Available at SSRN 3383359*, 2019.
[31] Tiancheng Shen, Jia Jia, Guangyao Shen, Fuli Feng, Xiangnan He, Huanbo Luan, Jie Tang, Thanassis Tiropanis, Tat Seng Chua, and Wendy Hall. Cross-domain depression detection via harvesting social media. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence, IJCAI 2018*, volume 2018-July, pages 1611–1617. International Joint Conferences on Artificial Intelligence, July 2018.
[32] Akkapon Wongkoblap, Miguel A Vadillo, and Vasa Curcin. Modeling depression symptoms from social network data through multiple instance learning. *AMIA Summits on Translational Science Proceedings*, 2019:44, 2019.
[33] Dawei Cong, Yanyan Zhao, Bing Qin, Yu Han, Murray Zhang, Alden Liu, and Nat Chen. Hierarchical attention based neural network for explainable recommendation. In *Proceedings of the 2019 on International Conference on Multimedia Retrieval*, pages 373–381, 2019.
[34] Hongxu Chen, Yicong Li, Xiangguo Sun, Guandong Xu, and Hongzhi Yin. Temporal meta-path guided explainable recommendation. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, pages 1056–1064, 2021.
[35] Hamad Zogan, Imran Razzak, Xianzhi Wang, Shoaib Jameel, and Guandong Xu. Explainable depression detection with multi-aspect features using a hybrid deep learning model on social media. *World Wide Web*, 25(1):281–304, 2022.
[36] Hamad Zogan, Imran Razzak, Shoaib Jameel, and Guandong Xu. Depressionnet: Learning multi-modalities with user post summarization for depression detection on social media. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 133–142, 2021.
[37] Jianlong Zhou, Hamad Zogan, Shuiqiao Yang, Shoaib Jameel, Guandong Xu, and Fang Chen. Detecting community depression dynamics due to

covid-19 pandemic in australia. *IEEE Transactions on Computational Social Systems*, 2021.

[38] Sandro Galea, Raina M Merchant, and Nicole Lurie. The mental health consequences of covid-19 and physical distancing: the need for prevention and early intervention. *JAMA internal medicine*, 180(6):817–818, 2020.

[39] Yeen Huang and Ning Zhao. Generalized anxiety disorder, depressive symptoms and sleep quality during covid-19 outbreak in china: a web-based cross-sectional survey. *Psychiatry research*, 288:112954, 2020.

[40] Chenxi Zhang, Lulu Yang, Shuai Liu, Simeng Ma, Ying Wang, Zhongxiang Cai, Hui Du, Ruiting Li, Lijun Kang, Meilei Su, et al. Survey of insomnia and related social psychological factors among medical staff involved in the 2019 novel coronavirus disease outbreak. *Frontiers in psychiatry*, 11:306, 2020.

[41] Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Russ R Salakhutdinov, and Quoc V Le. Xlnet: Generalized autoregressive pretraining for language understanding. *Advances in neural information processing systems*, 32, 2019.

[42] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*, 2019.

[43] Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using RNN encoder–decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1724–1734, Doha, Qatar, October 2014. Association for Computational Linguistics.

[44] Yoon Kim. Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1746–1751, Doha, Qatar, October 2014. Association for Computational Linguistics.

[45] Jeffrey Pennington, Richard Socher, and Christopher D Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543, 2014.

[46] Domenico Cucinotta and Maurizio Vanelli. Who declares covid-19 a pandemic. *Acta Bio Medica: Atenei Parmensis*, 91(1):157, 2020.