Post COVID-19 depression prediction using Twitter data

Eva Spirovska[1], Jovana Dobreva[1], Mary Lucas[3], Irena Vodenska[2], Lubomir Chitkushev[3] and Dimitar Trajanov[1,3]
[1] *Faculty of Computer Science and Engineering, Ss. Cyril and Methodius University, Skopje, N. Macedonia*
[2] *Administrative Sciences Department, Boston University, Metropolitan College, USA*
[3] *Department of Computer Science, Boston University, Metropolitan College, USA*

eva.spirovska@students.finki.ukim.mk, jovana.dobreva@finki.ukim.mk, mmlucas@bu.edu, vodenska@bu.edu, ltc@bu.edu, dimitar.trajanov@finki.ukim.mk

***Abstract*—This study aims to investigate the prevalence of Post COVID-19 depression by collecting, preprocessing, and analyzing English-language tweets using several natural language processing (NLP) models. Our primary objective is to identify depression-related tweets and develop a machine learning (ML) model for depression prediction. Two datasets are employed for this research: the first is a publicly available depression dataset from Kaggle, and the second is a long covid dataset obtained from Twitter between April 2020 and April 2022. By leveraging NLP techniques and ML algorithms, we analyze these datasets to gain insights into the pandemic's impact on mental health and identify key features associated with depression. Although the chosen classification model had promising results, it still misclassified certain data, prompting the incorporation of Twitter Account classification. Consequently, this integration resulted in tweets being classified more accurately.**

***Index Terms*—natural language processing, transformers, Twitter, mental health, depression**

I. Introduction

The COVID-19 pandemic has caused unprecedented global disruption, resulting in substantial loss of life and posing a significant threat to public health. In an effort to curb the virus's spread, governments and health organizations have enforced strict measures such as social distancing and isolation. Unfortunately, these measures have contributed to a surge in mental health issues, including anxiety, stress, worry, fear, and depression, which affected millions of people worldwide.

In these challenging times, social media has become an essential tool for communication and information dissemination about the outbreak [1]. Platforms like Twitter have been widely used to share information about the virus and allow individuals to express their thoughts and emotions about coping with the situation [2]. Research indicates that social media usage has increased during the pandemic [3], resulting in a considerable surge in activity. Consequently, social media platforms have evolved into valuable resources for identifying mental health concerns, including depression.

Analyzing Twitter users' responses enables us to gain a better understanding of the pandemic's impact on mental health and develop interventions to support those facing challenges. This research aims to create a model that can effectively address depression by analyzing data collected from Twitter. The proposed model will employ natural language processing (NLP) techniques and machine learning algorithms to examine tweets and identify depression indicators.

NLP, a subfield of artificial intelligence, is a crucial tool for understanding and analyzing human language. It combines computational techniques to process, comprehend, and apply human language to a diverse array of applications. Rapid advancements in NLP have facilitated the implementation of advanced deep neural networks, allowing for the extraction of relevant patterns from large text corpora. The significance of NLP extends to various domains, including healthcare [4] and pharmacology [5] that are closely related to our research topic.

Data science is a highly multidisciplinary field that uses complex algorithms to analyze the vast amount of data generated across different domains. It integrates principles from multiple disciplines, such as machine learning, mathematics, and statistics [6]. By combining NLP with data science methodologies, the proposed model can effectively leverage the potential of these complementary disciplines to identify and monitor depression indicators in Twitter users' language during the pandemic.

The remainder of this paper is organized as follows: Section 2 provides a comprehensive literature overview, discussing relevant studies in the fields of mental health, social media, and natural language processing. In Section 3, we describe the data collection process, including the sources and preprocessing of Twitter data used in this study. Section 4 outlines the methodology employed, detailing the natural language processing techniques and machine learning algorithms used to analyze tweets and identify depression indicators. In Section 5, we present the results of our analysis, highlighting the model's performance and the key features identified in depression classification. Finally, in Section 6, we draw conclusions based on our findings.

II. Literature overview

One of the most significant advancements in NLP is the development of transformers. Transformers, a relatively new type of architecture in deep learning, use the attention mechanism to weigh the importance of each portion of the input data separately [7]. BERT [8] and RoBERTa [9] are two powerful transformer-based models that have been built in recent years and have been widely used in natural language processing.

In the study conducted by Zhang et al. [10], the authors developed models for detecting and monitoring depression levels during the COVID-19 pandemic. They identified 2,575 distinct users with depression based on an analysis of their past tweets. Following this, they trained transformer-based depression classification models using the collected dataset. The results demonstrated that their fusion model achieved an accuracy of 78.9% on a test set comprising 446 individuals. Crucial features for depression classification were identified, including conscientiousness, neuroticism, first-person pronouns, discussions involving biological processes such as eating and sleeping, power-related conversations, and expressions of sadness.

A recent study [11] explores the use of Twitter to investigate depression and anxiety-related language during the COVID-19 pandemic. It highlights the significance of Twitter as an available data source for studying population mental health. To determine the frequency of depression, the researchers collected and analyzed 18 million tweets from Twitter users in seven major US cities during the stay-at-home period. They utilized machine learning-based language prediction models for anxiety and depression and concluded that depression and anxiety were frequently expressed on Twitter during this time, with an increase in negative emotions and a decrease in positive emotions. In addition, the study found a correlation between the trend of depression and national COVID-19 case trends and suggested social media as a potential tool for monitoring mental health.

Anwar and Ghosh introduce a method in their research paper titled "Depression intensity estimation via social media: a deep learning approach" [12] that utilizes social media data and deep learning techniques to estimate depression intensity. The approach involves extracting meaningful features from Twitter text data and training a small LSTM network to predict depression intensity. The authors found that depressed users frequently use negative words, post late at night, use personal pronouns frequently, and occasionally share personal events.

By employing sentiment analysis on Twitter posts, a study found that there was a significant increase in the use of depressive language during the COVID-19 pandemic in comparison to the pre-pandemic period [13]. In order to track public depression, the authors of this research analyzed tweets related to COVID-19. The study's results showed a significant increase in the depressive language during the pandemic, indicating that the COVID-19 pandemic has had a significant impact on the mental well-being of the general public.

Analyzing the emotions and attitudes expressed by individuals on social media has been suggested in [14] as a useful method for authorities to identify individuals who may be experiencing depression or be at risk for suicide. In another recent publication conducted by Angskun, Tipprasert, and Angskun (2022) [15], big data analytics was used to detect depression in real time. The researchers proposed a framework that involved the collection and analysis of data from social media platforms in identifying potential indicators of depression, including linguistic markers and behavioral patterns.

There have been several other studies that aimed to identify the primary topics posted by social media users related to the COVID-19 pandemic and applied sentiment analysis to detect the spread of anxiety and depression as a result of the outbreaks [16] [17].

III. Data

Data analysis is the process of obtaining, cleaning, transforming, and modeling data to extract essential information for a variety of decision-making processes. In this section, the whole process of data extraction, analysis, and preparation is explained in more detail.

For the purpose of this study, we use three types of datasets TableI. The first dataset [1] is a publicly available dataset from Kaggle which contains tweets related to depression from the social media platform Twitter. In addition, each tweet has a corresponding label indicating whether or not a particular tweet contains signs of depression.

The second dataset is collected with the help of the snscrape [18] Python library. This dataset also contains data obtained from Twitter, but unlike the previous one, all the tweets are related to the Post COVID-19 condition and have no labels [2]. Several attributes are present within the Post COVID-19 dataset, including *tweetUrl, tweetID, content, userName, userID, userLocation, isUserVerified, replyCount, retweetCount, likeCount, quoteCount, language, inReplyToTweetID, tweetConversationID, sourceLabel,* geographic information, etc. However, the focus of this study is on analyzing the text of a particular tweet. Thus, the other attributes present within the dataset are not being taken into account, and only the 'content' column is being considered for further analysis. In terms of language, the majority of tweets - up to 85%, are written in English. Because this study primarily focuses on tweets in the English language, only a small percentage of data is lost. A total of around 1.5 million tweets were analyzed in the predefined period between April 2021 and April 2022. Both of the previously mentioned datasets are used in the depression prediction phase.

For the Twitter account classification phase, an additional dataset is being collected. To gather all distinct Twitter accounts, the initial step is to retrieve the username of the author for each tweet included in the Post-COVID-19 dataset. Then, we utilize the Twitter data scraping tool - Twint [19] in order to scrape additional information for each respectively. After the completion of the scraping process, the resulting dataset for each includes the following information: id, name, username,

[1] https://www.kaggle.com/code/albertobellardini/depression-twitter/data
[2] https://www.who.int/europe/news-room/fact-sheets/item/post-covid-19-condition.

bio, location, URL, join date, join time, tweets, following, followers, likes, media, private, verified, account image URL, and background image. During the Twitter classification process, each of the features is taken into and utilized in the prediction phase.

TABLE I
DATASETS DESCRIPTION

	Kaggle dataset	*Post-COVID-19 dataset*	*Twitter account dataset*
Attributes	tweet content	tweet content	username, name, bio, location, name, URL, tweets, following, followers, likes and media
Labels	binary label (depressive tweet, non-depressive tweet)	no label	binary label (organizational account, personal account)
Removed data	/	non-English tweets	/
Data splitting	train set - 70%, val set - 15%, test set - 15%	dataset utilized for prediction	dataset utilized for prediction

A. Data preprocessing

Preprocessing data is crucial in the data analysis process as it leads to improved models and predictions. In NLP [20], data cleaning and tokenization are necessary steps to process text effectively for accurate interpretation.

In this study, we preprocess both the Post-COVID-19 and Kaggle datasets in the same way. The raw textual data from Twitter is typically unstructured and unorganized, and tweets can have up to 140 characters and include various elements like hashtags, mentions, URLs, and emojis in addition to the text. Considering that we are focusing only on word phrases, we must exclude all the unnecessary mentions, emojis, URL links, numbers, smileys, and hashtags. Therefore, we use the preprocessor library for Twitter data, developed by Said Özcan [21], in order to extract only relevant word phrases. This NLTK tool is used to remove stop words, punctuation, and lemmatization. Stop words are frequently used words that do not contain much useful information and may be eliminated without causing significant information loss. Sometimes, eliminating stop words may even enhance a natural language processing model's understanding and precision. Similar to stop words, punctuation can be removed without significantly affecting the meaning of the text, resulting in a text that comprises only words. With the removal of stop words and punctuation, we can improve the model accuracy, while lemmatization can standardize the text by simplifying words to their basic components. As a final step, we remove all the rows with missing tweet values. The final preprocessed dataset contains the original tweet, the cleaned tweet, and all URLs, mentions, hashtags, and emoticons placed in a separate column. As for the Twitter dataset, we preprocess it by only removing stop words, punctuation, and numbers.

B. Dealing with data imbalance

Imbalanced data is a common issue in classification problems, where the target class has an unequal distribution of observations. The problem arises when the classifier is biased toward the majority class due to a disproportionately high number of records in that class. To address this problem, oversampling, and undersampling techniques are used. Undersampling involves eliminating records from the majority class until it matches the number of records in the minority class while oversampling duplicates data from the minority class until it equals the number of records in the majority class. In this study, we apply these techniques to both the Post COVID-19 and depression datasets from Kaggle. As a result, from each dataset, two additional datasets are created, one using the undersampling technique and the other using the oversampling technique.

IV. METHODOLOGY

The entire approach to the models used for the purpose of this research will be explained thoroughly in this section.

A. Models used for depression prediction

We tried several distinct models and chose the model with the best accuracy for the final analysis. Since the Post COVID-19 data is not labeled, it is not feasible to evaluate the accuracy of the model with this data. Therefore, we use the labeled depression dataset from Kaggle. As described in the data section, this dataset contains tweets that are categorized as depressing (Label **1**) or not depressing (Label **0**), based on their content. The attribute that serves as an indicator of whether a tweet displays indications of depression or not is defined as the class attribute.

In this study, three models were implemented, including an LSTM neural network, as well as two transformer-based pretrained models, RoBERTa and DistilBERT. Further elaboration on their architecture is presented below.

*a) **LSTM Model:*** LSTM stands for Long Short Term Memory and represents a form of recurrent neural network [22]. These types of neural networks are considered to perform better in terms of memory than conventional recurrent neural networks since they are great at remembering certain patterns. They are very useful when working with Natural Language Processing. The LSTM model can determine the true meaning of the input text and provide the most accurate output class if suitable layers of embedding are used. The architecture of the LSTM model we use to forecast the corresponding issue of classifying depressive tweets is explained below.

To begin the analysis of the textual data, Word2Vec [23] is utilized to convert each word in the text into a corresponding numerical vector representation. This technique ensures that words with similar meanings have similar vector representations.

Subsequently, a sequential model is created and expanded with additional layers. The initial layer added to the model is a 32-neuron LSTM layer, followed by a dense layer as the output layer. The output layer has two cells representing the two classes (positive and negative), with a Softmax function set as the activation function.

The model is compiled using the Adam optimizer and categorical cross-entropy as the loss function. During the training phase, the model is trained and tested on all three types of datasets: imbalanced, balanced dataset with oversampling, and balanced dataset with undersampling. The training process is repeated ten times (epochs) to improve the model's accuracy.

b) ***DistilBERT Model:*** DistilBERT, introduced in [24], is a transformer-based model that is much faster and simpler than BERT. The overall architecture of DistilBERT is very similar to the BERT model described in [8]. DistilBERT is pretrained on the same data as BERT and employs the distillation technique that represents a compression strategy for training a smaller model to simulate the behavior of a bigger model. As described in the source paper, DistilBERT's 66 million parameters make it 40% smaller and 60% quicker than the BERT-base model, while keeping more than 95% of the BERT's efficiency. The model may be applied for either masked language modeling or sentiment analysis, but its primary goal is to be fine-tuned on a downstream task. We chose this model mostly because of its lower memory and its higher speed, and also because of its great performance when properly fine-tuned.

For the purpose of the study, the DistilBERT model is fine-tuned on the data that we use for our classification task. We use the Kaggle depression dataset to evaluate the performance of this model by classifying whether or not a tweet contains signs of depression. This problem represents a binary classification problem. At the beginning, it is essential to create a tokenizer object that is relevant to the model we use, because various pre-trained models use different ways to tokenize textual data. We then provide the model name to the ".from_pretrained()" method of the DistilBertTokenizerast class in order to obtain the tokenizer that is utilized by distilbertbase-uncased.

After creating the tokenizer object, we may proceed to utilize the proper tokenizer method to encode the training, validation, and test sets in batches. Following the tokenization technique, we import distilbert-base-uncased from the Hugging Face library to initialize the base model. The base DistilBERT model is used as a foundation, and then, on top of it, additional layers are added to fit our classification problem. By adding those extra layers, the task-specific knowledge of what distinguishes a tweet as depressed or non-depressed can be represented. On top of the underlying model, we add only one dense output layer with sigmoid activation. We use the Adam optimizer while compiling the model, with the learning rate parameter set to 5e-5, and the focal loss function used as a loss function. Regarding the training phase, only 4 epochs are used with a batch size equal to 64. The remaining parameters are left at their default values. Once more, all three datasets are utilized to evaluate the model's output, just like with the LSTM network. The results are shown in the final subsection of this appendix.

c) ***RoBERTa Model:*** RoBERTa represents a pretrained model on English language data using masked language modeling (MLM) objectives [9]. This means that it is pretrained using only raw data without any labeling. It uses Masked Language Modeling (MLM), where the model masks 15% of the inputs in a phrase and forecasts the masked words. With this approach, the model learns the internal representation of the English language, and then it can be used to extract features for a specific downstream task. Since this model is case-sensitive, it distinguishes between the words, for instance, English and English.

The approach for the RoBERTa base model is similar to the previous one using the DistiBERT model. The tokenizer object is initially created in order to properly tokenize the textual data that we use. The tokenizer used by Roberta-base is thus obtained by passing the model name to the ".from_pretrained()" function of the RobertaTokenizerFast class. After successfully generating the tokenizer object, we use it to encode the training, validation, and test data. As previously discussed, an additional output layer is added on top of the basic model to match our classification problem for depression. In the compiling phase, we use the Adam optimizer and the focal loss function as the loss function and learning rate set to 5e-5. The batch size in the training phase is set to 64, and the number of epochs is set to 4, while the other parameters are kept at their default values.

B. Models used for Twitter account classification

To conduct a more detailed examination and improve the final results, it is essential to identify the type of Twitter account associated with each tweet. This model addresses the account classification problem by using only the account's metadata. Its purpose is to classify Twitter accounts either as individual or organizational accounts. As a result, each Twitter account is labeled with 1 if classified as an individual and with 0 if classified as an organizational account.

Long Short-Term Memory (LSTM) network is used to build the model for textual processing and a fully-connected neural network for numerical processing. This classifier model achieves an average of 97.4% accuracy under 7-fold cross-validation. These results show that account metadata is a competent resource for identifying account types with high accuracy. The detailed architecture of the pretrained model is described in [25].

V. Results

In this section, several classification metrics are used to measure the performance of each model. Additional approaches for improving the result are also discussed.

A. Evaluation metrics

Although there are various metrics for classification models, the machine learning metrics Accuracy and F1 score are used to evaluate the performance of the classification models used throughout this study. A classification model's accuracy is determined by measuring how many of its predictions are correct as a proportion of the total predictions made. This metric is particularly useful when the classification model is working with data where the classes are evenly distributed. However, accuracy is no longer a relevant metric when dealing with imbalanced data. Another approach for addressing class imbalance issues is to use improved accuracy measures, like the F1 score metric. F1 score is composed of two other performance metrics, Precision and Recall. It combines the two metrics into a single metric and measures the model's performance. Due to the fact that in this research, we work with balanced and imbalanced datasets, both Accuracy and F1 score metrics are used as evaluation metrics.

B. Choosing the best model

We use the Kaggle dataset with depression tweets in order to determine the most effective model. For each model, three results are obtained: one from the original dataset, one from the balanced dataset using the oversampling approach, and one from the balanced dataset using the undersampling approach. Table II and Table III summarize the accuracy and binary F1 score achieved for each of the models.

TABLE II
F1 Score of the Models

Model	F1 score Imbalanced datset	F1 score Oversampled dataset	F1 score Undersampled dataset
LSTM	0.9714	0.9848	0.9556
DistilBERT	0.9830	0.9979	**0.9720**
RoBERTa	**0.9945**	**0.9991**	0.9692

TABLE III
Accuracy of the Models

Model	Accuracy Imbalanced dataset	Accuracy Oversampled dataset	Accuracy Undersampled datset
LSTM	0.9870	0.9849	0.9568
DistilBERT	**0.9920**	0.9979	**0.9726**
RoBERTa	0.9916	**0.9991**	0.9680

According to the outcomes, even though all of them are very similar, we can conclude that the LSTM model has the lowest accuracy and F1 score, while the DistilBERT and RoBBERTa models have greater and almost identical accuracy and F1 score. This can be attributed to the fact that DistilBERT and RoBERTa are transformer-based models that capture long-range dependencies, encode richer contextual information, and generalize well due to their pre-training on massive amounts of text data, leading to better accuracy on downstream NLP tasks. We can also note that the accuracy and F1 score are slightly higher when using a balanced, oversampled dataset. As a result, the RoBERTa model, in combination with a balanced oversampled dataset, is chosen as the best model for further analysis.

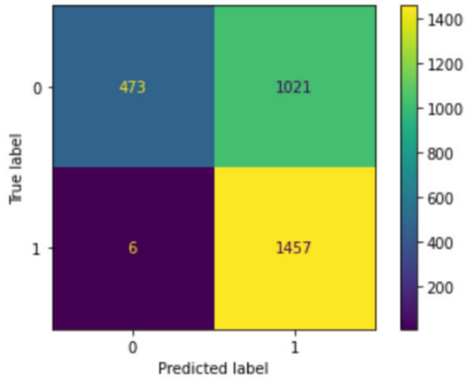

Fig. 1. Comparison between predicted labels and keyword search labels.

C. Final analysis and results

Considering that the RoBERTa model produces slightly better results than the other models, we chose this model for further analysis. We train our model using the accurately labeled tweets from the Kaggle depression dataset. Then, we use the dataset with Post COVID-19 related tweets to forecast the presence of depression. Since it is a problem of binary classification, a tweet can be classified as depressed with a value of 1 or not depressed with a value of 0.

Moreover, the test dataset is labeled using the keyword search technique. Several phrases that signify depression are used for this form of labeling, including the words 'depression', 'depressed', 'sad', 'moody', 'tearful', 'mental health', 'sorrow', 'hopeless', 'melancholy', 'misery', 'unhappy', 'despair' and 'melancholia'. After the successful labeling with keyword search, these labels are compared to those generated by the classification model. From the obtained results, it can be noted that there is a considerable difference in the outcomes. The majority of tweets classified as depressive using the RoBERTa model were identified as non-depressive using the keyword search labeling. The precision score is notably low, according to the results shown in Figure 1 and Table IV. Since the efficiency of the RoBERTa model has already been proven, it is apparent that the keyword search classification is not completely accurate.

Additionally, after a thorough analysis of the tweets, we observe that there are cases where the model identifies particular tweets as depressive, despite the absence of any evidence of depression. We also noticed that the falsely positive classification frequently occurs with tweets published by organizations, groups, etc. As an example, the tweet "@TorontoStar And to think nearly 60% will end up with #LongCOVID-19 as per @GovCanHealth statistics. This could be preventable." does not necessarily indicate signs of depression. It expresses concern and frustration about a particular statistic related to long COVID, a medical condition caused by the COVID-19 virus.

TABLE IV
SCORE OF THE COMPARISON BETWEEN PREDICTED LABELS AND KEYWORD SEARCH LABELS

	Precision	*Recall*	*F1 score*	*Support*
0	0.99	0.32	0.48	1494
1	0.59	1.0	0.74	1463
Accuracy			0.65	2957
Macro avg	0.79	0.66	0.61	2957
Weighted avg	0.79	0.65	0.61	2957

The writer acknowledges the high percentage of individuals who may develop long COVID-19 and hopes that this outcome could have been prevented. Although the use of the sad face emoticon suggests a negative emotional reaction, it alone does not necessarily indicate depression. The following Twitter post "@KristieMAllsopp How do we protect the mental health of the children that have already had #COVID-19 and are terrified of getting it again? Parents need the information to be able to make informed decisions, we don't want families to share our stories. #LongCovid" is another example that does not show any signs of depression. Instead, it expresses concern and a desire to protect the mental health of children who have already had COVID-19 and may be experiencing fear and anxiety about getting it again. While the topic is serious and may involve mental health concerns, the sentence itself does not include any clear indications of depression. Another relevant tweet that can serve as an example is the tweet, "Our members are rising. Please download our awareness pack. Help local schools, friends, and family by sharing our knowledge. We don't want new members, but as long as there is COVID-19, there will sadly be Long COVID-19 Kids.". The main objective of this tweet is to raise awareness about long COVID-19 and motivate people to participate in helping those affected by it. The language used is proactive, aiming to inspire people to take action. There is no apparent indication that the author is experiencing depression or feeling low.

Therefore, we conducted an additional classification of Twitter accounts, which resulted in the exclusion of tweets written by organizational accounts. The username of each tweet in the Post COVID-19-related dataset is employed to extract relevant information about the corresponding Twitter account. Following the gathering of data for each profile, the model introduced in [25] is used to classify them as individual or organizational accounts. Following the classification, all tweets created by Twitter accounts classified as an organization was excluded from the Post COVID-19-related dataset. This results in a decrease in the number of tweets classified as false positives.

VI. Conclusion

The objective of the classification problem we addressed in this paper was to accurately forecast depression in text. Twitter data related to COVID-19 was used during the period when the virus was most prevalent. We used three different models for the classification and fine-tuned them for our specific task. Among these models, the RoBERTa model, when combined with a balanced oversampled dataset, was chosen as the best choice for further analysis. Despite the fact that the classification model achieved good results, attaining an accuracy and F1 score of 0.9991, it was revealed that some data were incorrectly classified. As a result, Twitter account classification was also implemented, resulting in fewer falsely positive classified tweets, meaning that it was still possible to find tweets that were categorized as depressive even though they did not imply depression. Nevertheless, based on the models and techniques employed in this study, it can be concluded that even though the accuracy of identifying depression in the text is rather high, the result can still be improved using further approaches.

References

[1] J. Gao, P. Zheng, Y. Jia, H. Chen, Y. Mao, S. Chen, Y. Wang, H. Fu, and J. Dai, "Mental health problems and social media exposure during covid-19 outbreak," *Plos one*, vol. 15, no. 4, p. e0231924, 2020.

[2] M. Markovikj, J. Dobreva, M. Lucas, I. Vodenska, L. Chitkushev, and D. Trajanov, "Terminology and topics analysis of tweets related to the covid-19 pandemic," in *Terminology and Specialized Knowledge Representation: New Perspectives on User Needs*, Rijeka, Croatia, June 2022.

[3] M. Saud, M. Mashud, and R. Ida, "Usage of social media during the pandemic: Seeking support and awareness about covid-19 through social media platforms," *Journal of Public Affairs*, vol. 20, no. 4, p. e2417, 2020.

[4] D. Demner-Fushman, N. Elhadad, and C. Friedman, "Natural language processing for health-related texts," in *Biomedical Informatics: Computer Applications in Health Care and Biomedicine*. Springer, 2021, pp. 241–272.

[5] D. Trajanov, V. Trajkovski, M. Dimitrieva, J. Dobreva, M. Jovanovik, M. Klemen, A. Žagar, and M. Robnik-Šikonja, "Review of natural language processing in pharmacology," *Pharmacological Reviews*, 2023.

[6] I. Martinez, E. Viles, and I. G. Olaizola, "Data science methodologies: Current challenges and future approaches," *Big Data Research*, vol. 24, p. 100183, 2021.

[7] Y. Tay, M. Dehghani, D. Bahri, and D. Metzler, "Efficient transformers: A survey," *ACM Computing Surveys*, vol. 55, no. 6, pp. 1–28, 2022.

[8] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[9] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "Roberta: A robustly optimized bert pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.

[10] Y. Zhang, H. Lyu, Y. Liu, X. Zhang, Y. Wang, and J. Luo, "Monitoring depression trends on twitter during the covid-19 pandemic: Observational study," *JMIR infodemiology*, vol. 1, no. 1, p. e26769, 2021.

[11] D. Levanti, R. N. Monastero, M. Zamani, J. C. Eichstaedt, S. Giorgi, H. A. Schwartz, and J. R. Meliker, "Depression and anxiety on twitter during the covid-19 stay-at-home period in 7 major us cities," *AJPM focus*, vol. 2, no. 1, p. 100062, 2023.

[12] T. Anwar and S. Ghosh, "Depression intensity estimation via social media: a deep learning approach," 2021.

[13] M. K. Bashar, "Tracking public depression from tweets on covid-19 and its comparison with pre-pandemic time," in *2022 7th International Conference on Intelligent Informatics and Biomedical Science (ICIIBMS)*, vol. 7. IEEE, 2022, pp. 389–395.

[14] S. Sharma and S. Sharma, "Analyzing the depression and suicidal tendencies of people affected by covid-19's lockdown using sentiment analysis on social networking websites," *Journal of Statistics and Management Systems*, vol. 24, no. 1, pp. 115–133, 2021.

[15] J. Angskun, S. Tipprasert, and T. Angskun, "Big data analytics on social networks for real-time depression detection," *Journal of Big Data*, vol. 9, no. 1, p. 69, 2022.

[16] C. Cohrdes, S. Yenikent, J. Wu, B. Ghanem, M. Franco-Salvador, F. Vogelgesang *et al.*, "Indications of depressive symptoms during the covid-19 pandemic in germany: comparison of national survey and twitter data," *JMIR mental health*, vol. 8, no. 6, p. e27140, 2021.

[17] S. T. Sadasivuni and Y. Zhang, "Using gradient methods to predict twitter users' mental health with both covid-19 growth patterns and tweets," in *2020 IEEE International Conference on Humanized Computing and Communication with Artificial Intelligence (HCCAI)*. IEEE, 2020, pp. 65–66.

[18] "Justanotherarchivist snscrape: A social networking service scraper, version 0.4.3.20220106," *https://github.com/JustAnotherArchivist/snscrape*, 2022.

[19] C. Zacharias and F. Poldi, "Twint-twitter intelligence tool," 2017.

[20] E. Cambria and B. White, "Jumping nlp curves: A review of natural language processing research," *IEEE Computational intelligence magazine*, vol. 9, no. 2, pp. 48–57, 2014.

[21] E. Loper and S. Bird, "Nltk: The natural language toolkit," *arXiv preprint cs/0205028*, 2002.

[22] R. C. Staudemeyer and E. R. Morris, "Understanding lstm–a tutorial into long short-term memory recurrent neural networks," *arXiv preprint arXiv:1909.09586*, 2019.

[23] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.

[24] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter," *arXiv preprint arXiv:1910.01108*, 2019.

[25] Y. M. Cetinkaya, M. Gürlek, I. H. Toroslu, and P. Karagöz, "Twitter account classification using account metadata: organizationvs. individual," *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 30, no. 4, pp. 1404–1418, 2022.