Webology (ISSN: 1735-188X)
Volume 19, Number 3, 2022

Computing In Humanity: To Predict The Human Behaviors Over Social Media

Dr. Waleej Haider[1] , Engr. Muhammad Nadeem[2] , Sallar Khan[3] , Haris Ahmed[4] , Dr. Asad Abbasi[5] , Zainab Anwar[6]

[1,2,3,4]Sir Syed University of Engineering & Technology, Karachi, Pakistan.

[5]Benazir Bhutto Shaheed University Lyari, Karachi, Pakistan.

[6]Szabist University, Islamabad, Pakistan.

Abstract

Analyzing human behaviors using computer-based approaches is a new dimension of digital humanity. It integrates computing with healthcare, psychology, and social media.

Like other healthcare and medical-related problems, anxiety and depressive disorders are common in Pakistan. In the existing era, this disease is not only affecting people of all ages but the impact of depression is also witnessed on social media communications. On social media, anyone could not guess about the mental conditions of a user through his/ her tweets or comments. Especially, the prediction of depressed unknown users through his behavior on social media is a hard task. Computer science and the medical domain can eliminate such healthcare-related issues. In this paper, a system has been proposed to detect and analyze the behavior of a user on social media. The proposed system comprises of an algorithm that performs sentimental analysis of users' tweets and comments and a web-based platform to share best practices and success stories of users who have recovered from the disease. The system has been tested on real-time data obtained from tweeter using Application Programming Interface (API). Top trends have been focused to obtain the data and the proposed system successfully detected the users under depression. After working on the huge dataset, the proposed system will be a good contribution to humanity.

Keywords: Computing in humanity; Human behavior; depression mining

1. Introduction

The development and implementation of digital tools and procedures have exposed a significant impact on scholarly research. Large-scale digitization projects and hyperactive social media have brought social and historical texts, images, and other data that were previously difficult or impossible to reach into focus, raising new questions about how we conceive knowledge, think about human behaviors, and develop new epistemic practices. Some of the most important application areas for such digital humanities include cultural analysis, human behavior study and social interaction mining and depression detection.

In Pakistan, depression is a common cause of mental disease. Many individuals in Pakistan are uninformed about this issue. Depression isn't even considered an illness; rather, it's regarded as a dreadful disease. What they don't realize is that depression may be fatal as well. So, the majority of our society suffers from poor mental health and the fundamental problem is that most of them are unaware of it [1] [2] [3]. As a result, it is critical for them to recognize whether they are disturbed or require medical assistance. If people are conscious of their mental state, it is easier for others to understand them and be nice to them rather than being harsh on them without knowing what they are going through. Therefore, the identification of a depressed person has become a vital challenge and needs proper consideration [4] [5] [6]. On the other side, the majority of health information systems are useful, but none of them address mental health issues.

In contrast with the real-life physical environments, social networks are also victims of such issues as well. If online social posts are considered, it is found that a normal post by the user is different as compared to the posts shared by a depressed person. Further, it is also observed that teens with depressive posts are prevalent in social interaction platforms. Due to the unconscious depression, the words used in their social media posts get bitter, and overall, their post reflects a poor healthy mind [7] [8] [9].

In this connection, it is very critical to understand their specific behaviors and habits. This research has a goal to make depression detection simple and to raise awareness among individuals that depression cannot be avoided; it must be recognized. This research targets to address the aforesaid problem by employing sentiment analysis on social media comments. The keywords used in social media posts or tweets on Twitter can reflect user choices but can the reader really identify if the person is having anxiety or any other disorders? Probably not, because you never know the intention of a person behind the post. For instance, if a person shares depressive tweets often for a month and then shares funny content and then again posts the sad content, does it mean he/she was sad for a month and now he has good mental health? No, it indicates that his mental health is fluctuating.

In order to investigate such contexts, this research work emphasizes on social media comments. By employing artificial intelligence, the research has applied machine learning-based algorithms on the fetched tweets from social media networks and articulates if the person is depressed or not. The purpose of this research is to digitalize the processes of detection. Moreover, this also helps in monitoring the mental health of a variety of users such as students at colleges, universities, and business professionals.

The rest of the paper is organized as follows:

Section 2 explores literature related to existing approaches used for the detection of human behavior, Section 3 discusses thr proposed methodology. Section 4 presents the results and discussions. Section 5 articulates conclusions future directions.

2. Literature Review

As per the World Health Organization indications, depression has become the fourth most common problem and it is also predicted that it will prevail in the coming future. Depression patients are found to be in a state of depressed mental condition and they lack positive emotions [3] [5] [10]. Further, such patients prefer to be alone rather than being in a group setting. In this connection, the followings are some of the most promising study findings:

The work in [11] provides PYSCOM that collaborates with mental health doctors, seasoned authors, and health journalists to create content that helps you understand and sort through your emotions. The work also offers a depression survey which is based on Ivan Goldberg's Depression Screening Test. The Goldberg test may refer to a variety of psychiatric evaluations used to assess mental health in general or as screening instruments for specific mental syndromes such as depression or bipolar disorder. Another work in [12] is focused on utilizing user activities while using the web. Based on users' social media and web activities, including engagement with pages and adverts, the artificial intelligence powering these platforms determines what users see. Further, the work also assists people in locating therapy, resources, and support. Through continuing education and trainings, the ADAA aims to improve patient care by supporting the use of evidence-based therapies and best practices across disciplines and expediting the diffusion of research into practice. Another work [13] uses the demographics of the subjects, including their age, gender, education, and marital status, which were determined using a researcher-created questionnaire. In contrast with other works, the research work in [14] argues the use of standardized measurement instruments such as surveys to produce prevalence numbers. Authors also reason that this type of data does not provide a consistent, trustworthy picture of depression rates, noting that while birth cohort studies show an increase in depression prevalence, longitudinal studies do not reflect this.

This research was conducted at Shahid Beheshti University in Tehran, Iran. The resultant of the study was used as the study's dataset. A convenient sampling strategy was used to screen and recruit the survey users. Moreover, the research work in [15] articulates the recognition of mental health signs via a dynamic monitoring application. This also utilizes adaptive algorithms such as straight insertion sort to help users discover and recognize depression. In the long run, it also creates day-to-day suggestions for each user, which helps improve mental activities. Additionally, the research work in [16] is also directed towards the use of mobile applications for mood detection. This application is a micro-diary-based application that also enables the use of different machine learning models for detection.

In the same direction, the research work in [17] delivers an emotional health assistant mobile application powered by artificial intelligence. YOUPER app employs artificial intelligence to personalize a variety of strategies, including Extra Trees Regressor and Select from Model. Regression Estimators can be employed after converting the moods to decimal numbers. It also employs CBT (Cognitive Behavioral Therapy). Moreover, the research work in [18] provides a community-based web application that allows you to talk about your mental health with others. It employs a community detection technique that is adaptive. To maintain high-priority items at the top of the feed by deploying a machine-learning system. A comparative analysis of the existing method and proposed system is shown in table 1 in which the effectiveness of the system has been highlighted.

Table 1: Comparative Analysis of Existing Approaches

Reference	Characteristics			
	Patient Condition Analysis	**Mo trac**	**Depression Level Analysis**	**Repo Genera**
[11]	✓	✓	×	×
[12]	✓	×	×	×
[13]	✓	×	×	✓
[15]	✓	✓	×	×
[16]	✓	✓	×	×
Proposed Method	✓	✓	✓	✓

3. Proposed Methodology

The detection of depression is not an easy task. Especially, if it is diagnosed, then analyzing its level or severity is important for effective management. Traditionally used methods of detecting depression need the existence of a patient in front of a doctor for the diagnosis of disease. To serve humanity, it is needed to propose methodologies for the detection of depression by analyzing users' comments and tweets on social media. The main objectives of this research are:

Our system helps people to detect whether they are depressed or not through the analysis of their tweets. The goal is to make depression detection easy and spread areness among people that you can't run away from your depression, it has to be identified.

- Analysis of tweets based on keywords using the proposed algorithm.
- Generation of blogs based on precautions.
- Report generation.
- Success story to motivate patients for prompt recovery.

The main components of the system are as follows:

3.1 Data acquisition:

The system obtains the data in two ways. After generating an account, the users start the session by login an account and choosing an option as 1) filling a pre-designed questionnaire on our system or 2) by using tweeter account and sending tweets on a particularly hot issue. To fetch the data from tweets it was not possible to get all the tweets shared on a particularly hot issue due to restrictions of the tweeter. Therefore, a tweeter developer account has been generated so that the tweets of specific persons could be obtained. For this purpose, we have applied for a developer account through the Twitter developer portal and when the account was created we got the API keys that authenticate our connection to Twitter so that we can collect the tweets of the relevant person(s). Python has been used for coding. The following code has been used to access the user accounts using an API.

```
authenticate=tweepy.OAuthHandler(CONSUMER_KEY,CONSUMER_SECRET)
authenticate.set_access_token(ACCESS_TOKEN,ACCESS_SECRET)
for tweet in posts [0:5]:
print (tweet.full_text)
```

Some hot issues have been selected to acquire the tweet data. In both cases (questionnaire and tweets), mixed data is obtained. A natural Language Procession (NLP) has been used to acquire the meaningful data from the tweets to be used by the algorithm. NLP extracts the aspects from the data. These aspects work as keywords that help in data analysis.

3.2 Data Preprocessing:

The acquired data could not be used by the algorithm in the existing status for sentimental analysis. This data needed to be cleaned. Therefore, first, we consider the fetched tweets and pre-process them to remove unwanted problematic characters, or special characters hashtags and hyperlinks which are unreadable and also not useful for the algorithm. Now the data is ready to be loaded to the algorithm for sentimental analysis.

3.3 Algorithm Implementation:

The proposed algorithm is a combination of codes consisting of libraries used for a sentimental analysis of a text dataset. We used TextBlob library in the algorithm which performs sentiment analysis on the tweets.

TextBlob is a library of python which is used in NLP for implementing sentiment analysis. Sentiment analysis is the process in which we determine the emotion of the writer that can either be positive, negative, or neutral. We get these negative and positive values from the two functions of TextBlob library. The first function is subjectivity and the second one is polarity. Polarity determines the values of tweets as -1 and 1. These values represent the sentences of tweets or comments as negative or positive for a better understanding of the algorithms. While subjectivity refers to the emotions in text and gives values 0 and 1 to the text. In this part of the algorithm, we have provided fetched tweets to TextBlob functions of subjectivity and polarity to perform

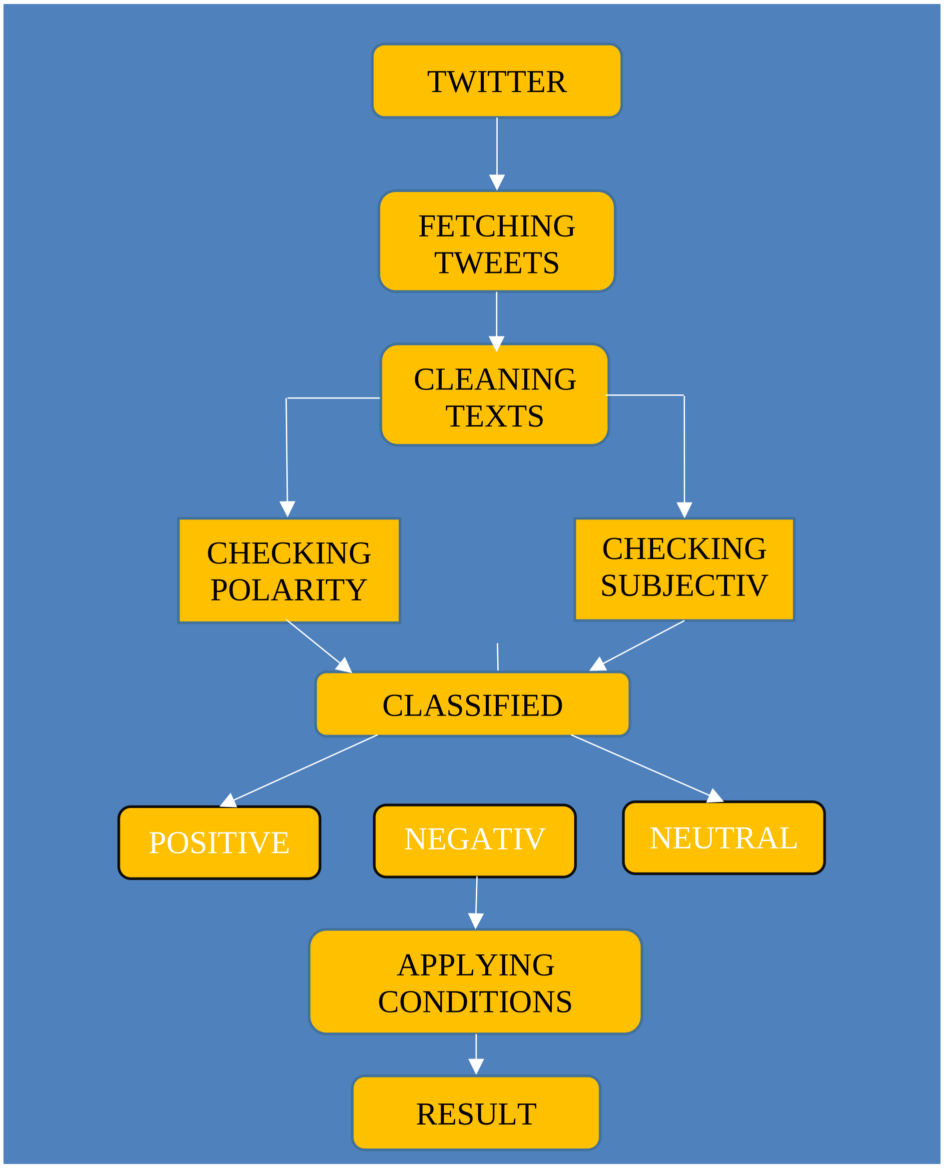

Figure 1: Working Flow of Proposed Algorithm

sentiment analysis and differentiate them as positive, negative, or neutral. This classification helps the system in taking accurate decisions as shown in Figure 1.

Now after classifying tweets the system has sorted them and checked their percentages according to their occurrence. First, the system has sorted tweets based on their polarity values and only printed positive tweets. Similarly, in the next step, the system has sorted tweets again and printed negative tweets. After classifying and sorting them the system calculated the percentages of tweets as positive, negative and neutral separately.

After analysis, all the tweets have their corresponding values. Accordingly, the values of polarity and subjectivity functions of TextBlob have also been computed. Now it becomes easy to judge the nature of tweets based on their polarity and subjectivity ratings. Finally, after classifying the tweets and finding the percentages of positive, negative, and neutral tweets the system have applied conditions on algorithm results and predicted the result regarding human behavior. At this stage, the algorithm has determined the mental state of the users based on specified conditions. Conditions are represented by the algorithm as if the negative tweets have the highest percentage or frequency as compared to both neutral and positive tweets then the user is diagnosed as depressed by the proposed algorithm.

3.4 Disease Management by using Human experiences:

Human behavior analysis using computer-based approaches could not be effective without the involvement of the human who has already experienced the situation and recovered. The proposed system encompasses a web-based platform for the sharing of problems of depressed users, remedies and possible solutions of problems and success stories of the users. A user can get benefit from the success story of another user who has recovered from depression. He can apply the same routine practices, exercises, and rehabilitation sessions to get rid of the disease and can share his

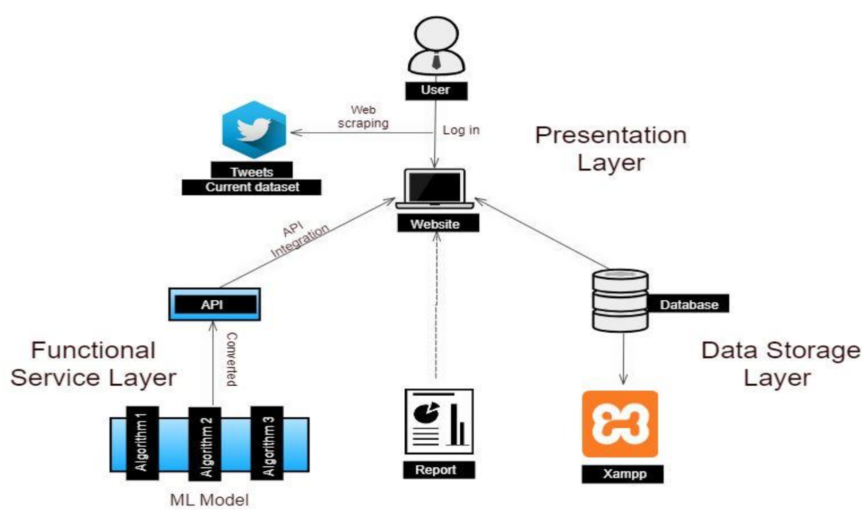

Figure 2: Web based Platform to Share user Experiences and Success Stories

stage-by-stage experience with other users.

The architecture of the web-based platform is shown in figure 2. The overall outline in the system architecture diagram is comprised of the entities includes a Machine Learning model, a database, Application Programming Interface (API) and a web application. The model is created at a functional service layer containing an algorithm. The front end is a website where the results are presented and generated in the form of an actual report.

The system obtained the user data from other social media sites using web scraping and stored it in a database for further use. Web scraping is a method of extracting required information from various websites using codes called bot or crawlers. An API is used to connect the systems with a functional service layer where a variety of ML algorithms are available for data processing. A cross-platform web service has been used or script writing in the used programming language. Reports can be generated after processing user data using machine learning algorithms.

```
C:\Users\salma\PycharmProjects\noApiAllcode\venv\Scripts\python.exe C:/Users/salma/PycharmProjects/noApiAllc
Enter Twitter Account Name:ASIMSBAJWA
This is a great decision as needs of Balochistan health care system specially for rural areas can never be m
RT @CathayPak: The Embassy highly appreciated @Asad_Umar and @AsimSBajwa met representatives from CPEC proje
RT @Asad_Umar: Held a meeting yesterday with all the companies working on coal to gas &amp; coal to liquid p
Condolences on the sad demise of highly respected former PM Zafarullah Jamali sb. Prayers for him and his fa
https://t.co/K0syl8gIcb here is the link for those of you who missed reading this article by the newly arriv
                                              Tweets
0  This is a great decision as needs of Balochist...
1  RT @CathayPak: The Embassy highly appreciated ...
2  RT @Asad_Umar: Held a meeting yesterday with a...
3  Condolences on the sad demise of highly respec...
4  https://t.co/K0syl8gIcb here is the link for t...
0     This is a great decision as needs of Balochist...
1     : The Embassy highly appreciated _Umar and  me...
2     _Umar: Held a meeting yesterday with all the c...
3     Condolences on the sad demise of highly respec...
4      here is the link for those of you who missed ...
```

Figure 3: Results of Obtained Tweets

4. Results and Discussions

The proposed model works on tweets and comments of social media users to analyze and predict human behavior. For this purpose, the TextBlob library uses its two functions i) subjectivity and ii) polarity to perform sentiment analysis and after that classify tweets as either positive, negative, or neutral. We have selected some hot trends on social media and found tweets of particular users on these trends. For example, first, we have used SIMBAJWA as a common trend of social media. The results of tweets on this trend are as follows:

The system receives the tweets in raw form and converted these into data frames to prepare the data for cleaning as shown in Figure 3. These data frames have been converted into corpus frames that represent the data by using the names of each text column to make data mining easy. Using corpus, an algorithm only focuses on specified text lines for mining by avoiding other paragraphs. Now the process of data cleaning starts in which the algorithm has removed irrelevant text from the tweets and counted the sentiments in the remaining tweets. The system has created new data frames that have tweets with specific mean values.

In the last step, the cleaned tweets have been classified by the algorithm as positive, negative and neutral and found the values in percentage (%). Here it can be seen in Figure 4 that the output of the final part of the code and results of the algorithm that displays the percentages of tweets as the algorithm have obtained 65% positive tweets 30% neutral and just 5% negative tweets

```
Positive Tweets Percentage
65.0
Negative Tweets Percentage
30.0
Neutral Tweets Percentage
5.0
======RESULTS======
Not Depressed
```

Figure 4: Classification of Tweets of a User

of a user on a selected topic that means the posts of this user showed no symptoms of depression. Based on these results the system predicts the mental state of the person as not depressed.

Finally, the system has created an API Here using flask which is a framework used to create rest and restful APIs as shown in Figure 5. In our case, we have created an API of our code that will be used to pass parameters into the algorithm through our web-based platform. The results are displayed on this platform.

This API has been integrated with our web-based platform. When a user enters his tweeter

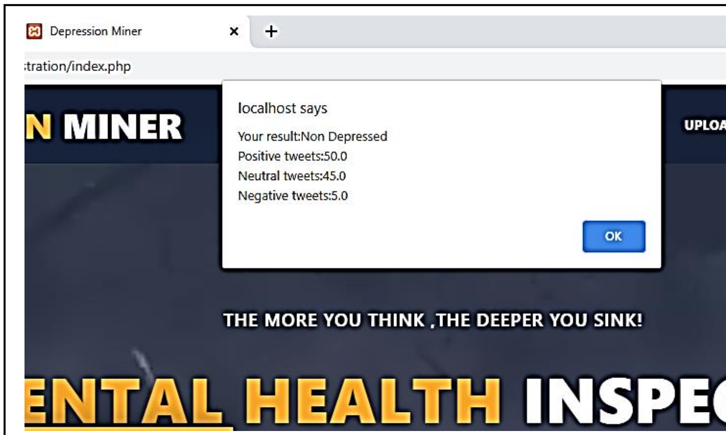

Figure 5: API Created by Proposed System

username through our website, the API displays the results of the user at the website on runtime.

It is important to compute the frequency of specific words used by a user in his/her tweets to find his mental status. If he used harsh slang words more than a threshold value, our system will consider him in a depressed condition and vise versa. For this purpose, the system has used word cloud to see the frequency of words used in the shared tweets with a graph so that we can find the most used words by the user which will help us in checking the accuracy of results of our algorithm as shown in Figure 6.

Word cloud is created by importing its library in code and then when we feed tweets in it the image, which is created, shows us the most frequently used words by the users in their tweets that helps us to analyze the overall sentiment of the tweets better. The system can work on analyzing tweets of multiple users on a particular topic simultaneously. Also, the proposed system can work on multiple hot topics to analyze human behavior using their tweets and comments shared on social media.

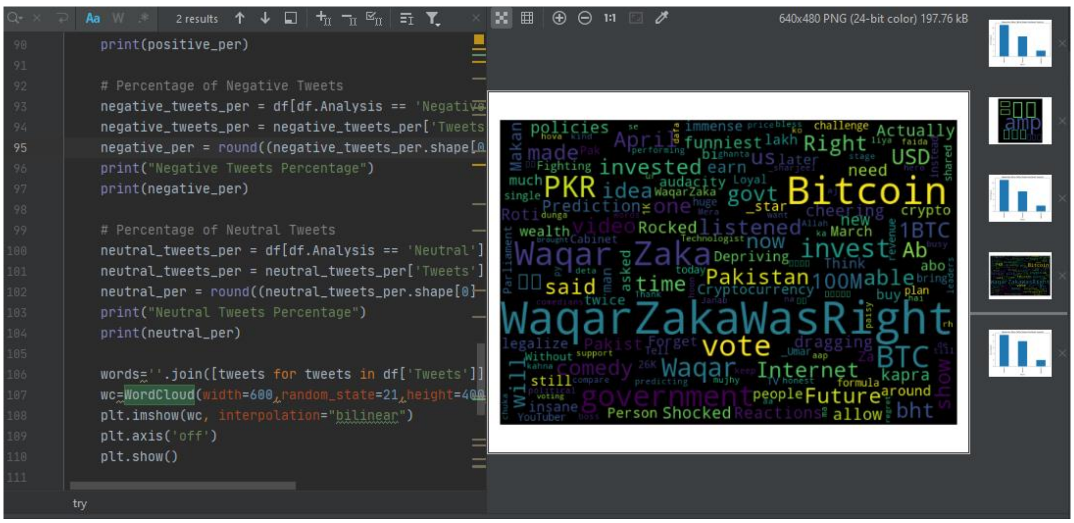

Figure 6: Word Cloud to Calculate the Frequency of Words in the Tweets

5. Conclusion and Future Directions:

The proposed work aims to provide a computer-based solution for humanity to analyze human behavior using their tweets and comments shared on social media. For this purpose, an algorithm using TextBlob library has been proposed for sentiment analysis and a web-based platform has been presented for sharing current conditions of depressed users, success stories of the users who recovered from the depression disease. The system will save the stage-wise records of depressed users while using best practices of success stories. Our system can effectively analyze human behavior with good accuracy and facilitates the users to adopt the best practices of successful users.

References:

[1] D. Kessler, O. Bennewith, G. Lewis, and D. Sharp, “Detection of depression and anxiety in primary care: Follow up study,” Br. Med. J., vol. 325, no. 7371, pp. 1016–1017, 2002, doi: 10.1136/bmj.325.7371.1016.

[2] S. Alghowinem, R. Goecke, M. Wagner, G. Parker, and M. Breakspear, “Eye movement

analysis for depression detection," 2013 IEEE Int. Conf. Image Process. ICIP 2013 - Proc., pp. 4220–4224, 2013, doi: 10.1109/ICIP.2013.6738869.

[3] M. R. Morales and R. Levitan, "SPEECH VS . TEXT : A COMPARATIVE ANALYSIS OF FEATURES FOR DEPRESSION DETECTION SYSTEMS Michelle Renee Morales Department of Linguistics Rivka Levitan Brooklyn College , CUNY Department of Computer Science," 2016 IEEE Spok. Lang. Technol. Work., pp. 136–143, 2016.

[4] X. Wang, C. Zhang, Y. Ji, L. Sun, L. Wu, and Z. Bao, "A depression detection model based on sentiment analysis in micro-blog social network," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 7867 LNAI, pp. 201–213, 2013, doi: 10.1007/978-3-642-40319-4_18.

[5] H. Cai et al., "A Pervasive Approach to EEG-Based Depression Detection," Complexity, vol. 2018, 2018, doi: 10.1155/2018/5238028.

[6] M. R. Islam, M. A. Kabir, A. Ahmed, A. R. M. Kamal, H. Wang, and A. Ulhaq, "Depression detection from social network data using machine learning techniques," Heal. Inf. Sci. Syst., vol. 6, no. 1, pp. 1–12, 2018, doi: 10.1007/s13755-018-0046-0.

[7] Y. Gong and C. Poellabauer, "Topic modeling based multi-modal depression detection," AVEC 2017 - Proc. 7th Annu. Work. Audio/Visual Emot. Challenge, co-located with MM 2017, pp. 69–76, 2017, doi: 10.1145/3133944.3133945.

[8] A. H. Orabi, P. Buddhitha, M. H. Orabi, and D. Inkpen, "Deep learning for depression detection of Twitter users," Proc. 5th Work. Comput. Linguist. Clin. Psychol. From Keyboard to Clin. CLPsych 2018 2018 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. NAACL-HTL 2018, pp. 88–97, 2018, doi: 10.18653/v1/w18-0609.

[9] S. G. Burdisso, M. Errecalde, and M. Montes-y-Gómez, "A text classification framework for simple and effective early depression detection over social media streams," Expert Syst. Appl., vol. 133, pp. 182–197, 2019, doi: 10.1016/j.eswa.2019.05.023.

[10] M. Deshpande and V. Rao, "Depression detection using emotion artificial intelligence," Proc. Int. Conf. Intell. Sustain. Syst. ICISS 2017, no. Iciss, pp. 858–862, 2018, doi: 10.1109/ISS1.2017.8389299.

[11] I. Goldberg, "No Title," 2020. https://www.psycom.net/editorial-policy-and-guidelines/.

[12] Z. Neumann, "ADDA," 2021. https://adaa.org/.

[13] B. Ajilchi and V. Nejati, "Executive functions in students with depression, anxiety, and stress symptoms," Basic Clin. Neurosci., vol. 8, no. 3, pp. 223–232, 2017, doi: 10.18869/nirp.bcn.8.3.223.

[14] H. P. Hamer and S. Clarke, "No Title," 2014. https://mentalhealth.org.nz/assets/Our-Work/policy-advocacy/depression-epidemic-.

[15] M. O'Leary, "No Title," 2020. https://onemindpsyberguide.org/apps/minddoc-app-review/#:~:text=Moodpath supports users in tracking,share with a health provider.

[16] S. McLane, "No Title," 2019. https://samclane.dev/Daylio-Analysis/.

[17] UKDiss, "No Title," 2019. [Online]. Available: https://ukdiss.com/examples/machine-learning-depression-screening.php#citethis.

[18] I. C. Education, “No Title,” 2020. https://www.ibm.com/cloud/learn/supervised-learning.