

Multi Model Machine Learning Approach for Predicting Mental Health Outcomes

Yash Narkhede¹, Aditya Nehete², Swapnil Sakpal³, Sanket Jaiswal⁴ and Gresha Bhatia⁵

¹⁻⁵VES Institute Of Technology, Mumbai, India

Email: {2020.yash.narkhede, 2020.aditya.nehete, 2020.swapnil.sakpal, 2020.sanket.jaiswal, gresha.bhatia}@ves.ac.in

Abstract— The proposed metric capacity unit methodology to classify patients' mental disorders based on their posts and relevant comments on Twitter is a unique and innovative approach to addressing the issue of leveraging machine learning algorithms for e-health. The aim was to investigate the effectiveness of different machine learning models for predicting mental health outcomes. Specifically, we compared the performance of a traditional single-model approach with a novel multimodal approach that combines multiple models to enhance predictive accuracy. Furthermore, the multimodal approach was particularly effective for predicting complex mental health outcomes that are difficult to capture with a single model. In addition, the study also used visualization techniques to provide explainability to the models used in the multimodal approach. This helped to understand how different models contribute to the overall prediction and identify which features were most important for predicting mental health outcomes. Overall, this unique approach has the potential to greatly impact the field of e-health and mental health, by leveraging the power of social media and machine learning algorithms to diagnose and address mental health issues in a more efficient and accessible manner.

Index Terms— Natural Language Processing (NLP), Mental Health, Multimodal Learning, BERTViz.

I. INTRODUCTION

Social media has become an integral part of our daily lives, with users posting and communicating an incredible amount of information every day. This connectivity has also made stress more visible and prevalent than ever before. [2]With platforms like Twitter, Reddit, and Facebook, researchers have access to a wealth of data to study the worries and stresses of people worldwide.

Machine learning (ML) algorithms have proven to be effective in the diagnosis and treatment of internal diseases. The pattern recognition capabilities of ML, combined with big data analysis, can provide sophisticated and accurate results. The objective of this research paper is to investigate the application of several machine learning algorithms - including Decision Trees, Support Vector Machines, Naïve Bayes, Logistic Regression, and k-nearest Neighbor classifiers - for the analysis of the mental health condition of a specific group of individuals, by studying their social media posts.

Our research aims to provide a more accessible and efficient way to diagnose and address mental health issues globally[4][5]. By leveraging the power of big data and deep learning techniques, we hope to develop a sophisticated model for diagnosing mental health issues based on social media data. This research has the potential to revolutionize the field of e-health and improve the lives of people suffering from mental health issues.[6]

II. OVERVIEW

A. Motivation

Social media platforms, such as Twitter, have become a major source of emotional expression for individuals around the world. By analyzing tweets, researchers can gain insights into the emotional states of people in different contexts, such as during a pandemic, natural disaster, or political event. Emotion prediction based on tweets can have practical applications in fields such as marketing, customer service, and public opinion analysis. For example, The ability to predict emotions can be advantageous for companies who wish to track customer attitudes towards their products, or for political campaigns who want to assess the public's perception of their candidate or proposed policies. Emotion prediction can be useful in mental health research and treatment. By analyzing the emotions expressed in tweets, researchers and clinicians can gain insights into the emotional states of individuals and potentially identify those who may be at risk for mental health problems. Predicting emotions based on tweets can also have broader implications for understanding human psychology and social behavior[7].

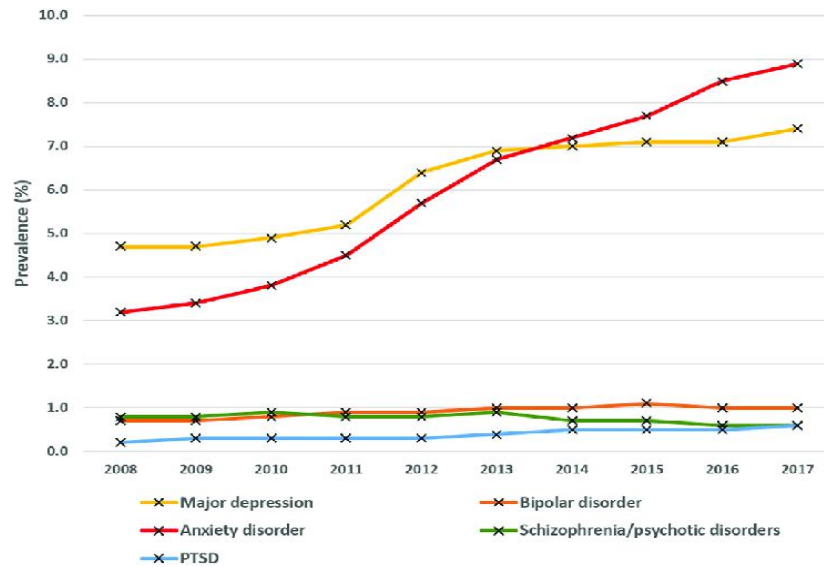


Fig 1. Cases Per Year

B. Problem Statement

Mental health conditions are a growing concern in today's society, affecting a significant portion of the population. Depression, [10] anxiety, and loneliness are among the most common mental health conditions that can have serious consequences if left untreated, including a negative impact on physical health, social relationships, and quality of life.[8] Social media has become an increasingly popular platform for communication and interaction, with billions of people using it every day. The vast amount of data generated by social media provides a unique opportunity to detect and predict mental health conditions in individuals, potentially enabling early intervention and prevention of these conditions. The goal is to develop a system that utilizes ML techniques to predict mental health conditions based on social media chats. Our system utilizes multimodal capabilities[13], incorporating data from multiple sources such as text, audio, and images, to enhance the accuracy and reliability of the predictions. This information is then used to provide audit prescriptions, including personalized recommendations for intervention and referral to mental health professionals if necessary.

C. Objectives

- To predict the emotion of a person based on social media and then predict if the person is suffering from any mental condition
- To check if proper dosage is given for a particular mental health condition
- Enhance the accuracy of the outcome by employing various techniques and strategies.
- An interface that is easy to use and navigate, which enhances the user experience.
- Improve the efficiency of prediction in this domain
- Visualizations in our system can help provide explainability to the models used in the multimodal approach

III. THEORY

Through an extensive review of IEEE papers cited in our references, we have come to the conclusion that detecting psychological stress in online posts is a highly sensitive application that requires the ability to interpret models accurately.[3]Through our use of emotion-based models, we have achieved results that are on par with the latest BERT model. [16]Additionally, we have developed an interactive visualization feature that improves the interpretability of our results. These unique features of our approach make it an innovative solution to the challenge of detecting psychological stress in online posts.

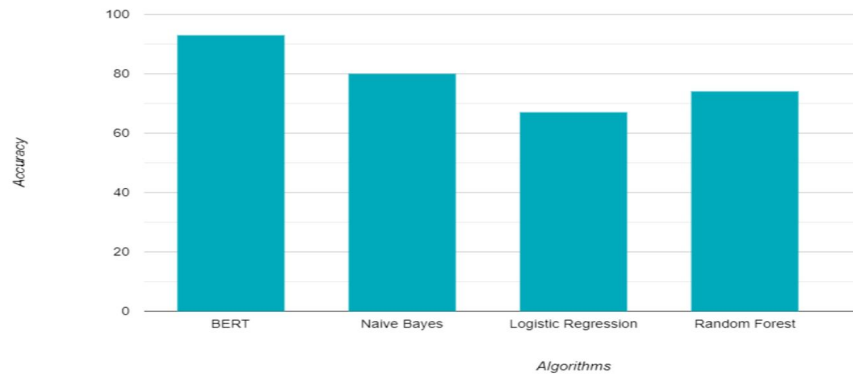


Fig 2. Accuracy Variation

A. Survey of Existing Systems

To create this system, we tested out BERT and DistillBert Models. This paradigm is one of the most well-known and frequently employed ones in the study of sentimental and mental disorders. To the best of our knowledge, no previous research project in this industry has attempted to undertake a comprehensive analysis of both sentiment and mental health disorders in a combined manner, according to our study of reference materials, and the majority of projects or businesses only use traditional sentiment analysis or only use social media data, including Twitter, to predict mental health disorders. After analyzing the existing literature, we found that [11] Naive Bayes, Logistic Regression, and Random Forest were the most popular models for sentiment and mental health disease prediction. However, it was noted that these models did not perform as well as the BERT model. ML models like Naive Bayes for sentiment analysis had an accuracy of 75-80%. [1] The accuracy of the Logistic Regression and Random Forest algorithms for predicting mental health disorders was 75.17 percent and 76.31 percent, respectively. BERT, which uses fine-tuning and hyperparameter tuning, had a slightly higher accuracy here 81.9 percent. [12] The adaptability of the BERT model stems from its ability to recognize intricate patterns in textual inputs and its ability to focus on certain tasks, and the fact that it has already been taught, making it a good place to start for many natural language processing jobs.

B. Limitations and Research Gap

Since the majority of the currently utilized systems employed dated and common techniques for their models, they are not particularly effective. After being run, these models provide a relative accuracy that is lower. The current model does not yield significant findings, having an average precision of only 71% [20]. Not only that, but other parameters, including the total loss, also show a significant decline. Our proposed system, in contrast to existing systems, utilizes new, faster transformer-based algorithms that are not commonly used in mental health prediction.[9] Additionally, our model includes multimodal capabilities, providing explainability and visualization that are lacking in many existing models [19]. Through our analysis of IEEE/Springer papers, we have identified flaws and drawbacks in existing research that our proposed system aims to address. By incorporating these advanced features, our system aims to provide more accurate and precise results, ultimately improving the overall effectiveness of mental health prediction and treatment.

C. Project Contribution

Team 1: Requirements and Planning

The problem statement was identified during the project's first phase, which also included careful requirement collection. One team member led the team in identifying the project goal of predicting mental health based on

social media data, while two other members conducted requirement gathering. The team worked together to prioritize and carefully collect the necessary requirements and made sure all fields were maintained during development. Good planning and record-keeping helped ensure that the software met the needs of the user and that accurate cost estimates could be made.

Team 2: Design and Implementation

The second phase of the project involved designing and implementing a machine learning-based system for predicting mental health based on social media data with prescription audit, using the BERT model for greater accuracy and a multimodal concept. Some team members conducted testing, while others assessed the system's effectiveness.

IV. BERT MODEL FOR PREDICTION AND VISUALISATION

A. Architecture

The Transformer serves as the foundation for BERT's architecture. The general structure of BERT is made up of a number of Transformer blocks, each of which is made up of an attention mechanism.

BERT was initially made available in base and big variations for both case-sensitive and case-insensitive input text. We utilized the 110M-parameter bert-base-uncased.

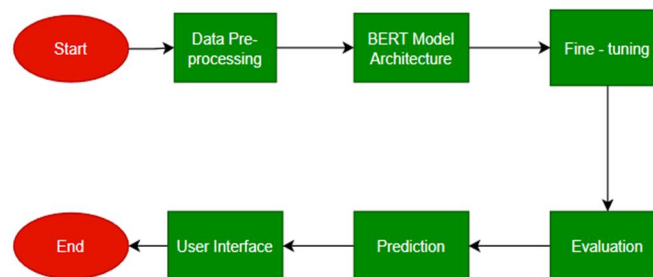


Fig 3. Modular Diagram

The process of building a mental health prediction system using social media data involves several key steps. Firstly, the preprocessed and cleaned data is given as input to a BERT model for predicting the emotional state of a person. To improve the accuracy of the model, attention layers and attention weights are used to focus on the most important parts of the text. The BERT model outputs the predicted emotional state along with the original tweet text. Secondly, the output from the emotional state prediction model, along with the original tweet text, is given as input to another BERT model for predicting the mental health status of the user. This model uses multi-class classification to predict the presence or absence of various mental health disorders based on the user's tweets. Similar to the emotional state prediction model, attention layers and attention weights are used to improve the accuracy of the mental health prediction model. To make the model more interpretable and gain insights into its inner workings, we utilized the BertViz tool, which allows for a detailed analysis of the attention weights and important text passages that contribute to the model's predictions. This helped us gain a deeper understanding of how the model works and identify areas where further optimization may be necessary.

B. Algorithm and Process Design

To design an effective mental health prediction model, it is crucial to follow a well-defined algorithm and process. [15] Initially, a comprehensive dataset is collected from various social media platforms, which is then subjected to rigorous pre-processing techniques such as data cleaning and augmentation to extract relevant features and information while eliminating noisy data. [14] Subsequently, a suitable NLP algorithm such as BERT is chosen, and the pre-processed data is divided into separate training and testing datasets to evaluate the model's performance accurately. The processed data is then fed as input to the BERT model to predict the emotional state of a person with high precision and accuracy. Attention layers and attention weights are used to focus on the most important parts of the text, thus making the model more accurate. The output obtained from the emotional state prediction model along with the original tweet text is given as input to another BERT model for the prediction of the mental health of the user. Multi-class classification is used to classify the user's mental health status. The algorithm

selection process involves training the algorithm on a designated training dataset, fine-tuning the hyperparameters to improve model performance, and evaluating the trained model on a separate testing dataset using performance metrics such as precision, recall, F1-score, and accuracy. Depending on the evaluation results, the model can be further refined by fine-tuning hyperparameters or selecting a different algorithm altogether. Once the model is trained, evaluated, and optimized, it can be utilized to predict outcomes on unseen data.

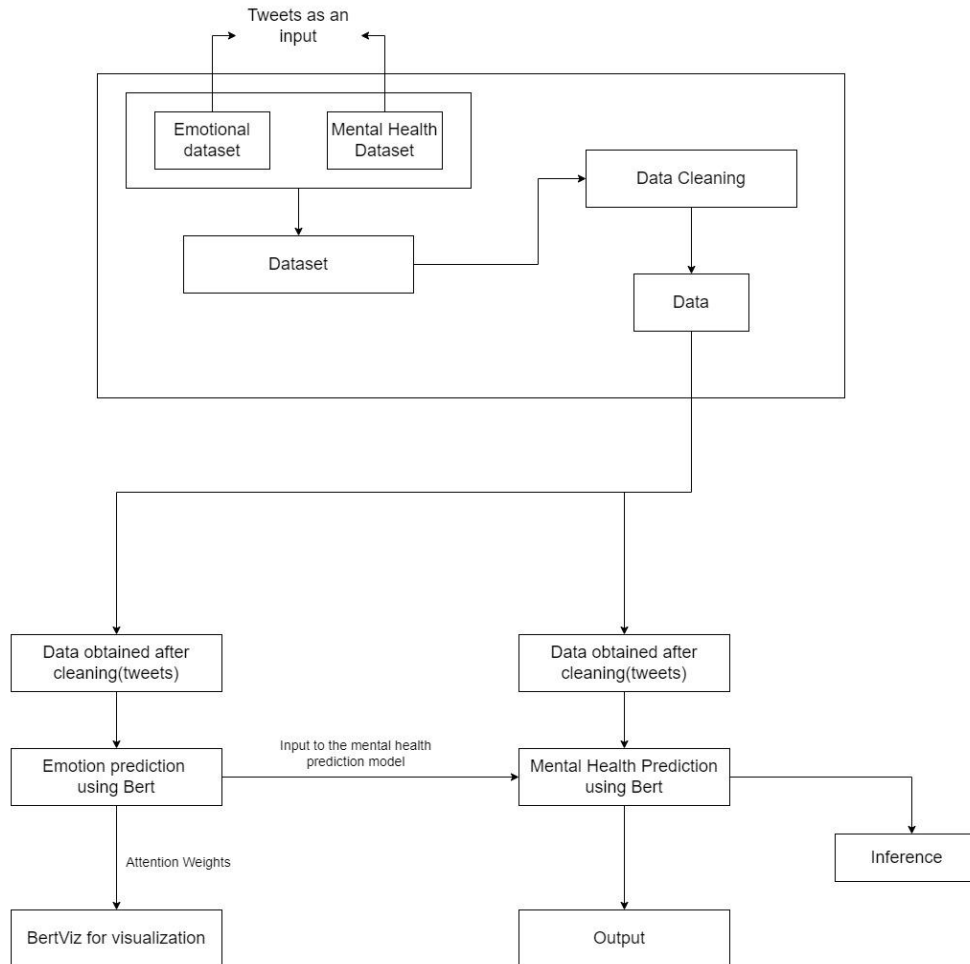


Fig 4. Flow Diagram

V. RESULTS AND EVALUATIONS

A. Verification and Validation

During the development of our model, we performed a thorough evaluation to ensure that it met all of the necessary requirements. To achieve this, we divided the dataset into separate training, validation, and testing sets. By doing so, we were able to evaluate the model's performance in a suitable environment and verify that it effectively accomplished its intended objectives. The results obtained from these sets were analyzed, and the conclusions derived are presented in the following figures.

The incorporation of the BERT model has demonstrated significant success in anticipating the emotional state and psychological health condition of users. This groundbreaking technology provides an improved, user-friendly, and low-cost solution for consultations with patients. Our ultimate solution comprises a sentiment and mental health predictor that is presented on a web application powered by the BERT model. Additionally, the use of optimized algorithms has helped to reduce loss. With further development of the model and expansion of the dataset, it may be possible to predict the appropriate dosage for various diseases, leading to even greater improvements in the model's functionality. Fig shows the sentiment detected, Fig shows the predicted mental health disease while Fig shows whether, for the particular disease, proper medication is given to the user or not.

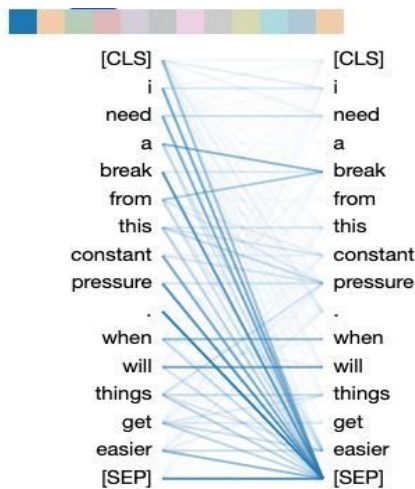


Fig 5. Visualization using BERTViz

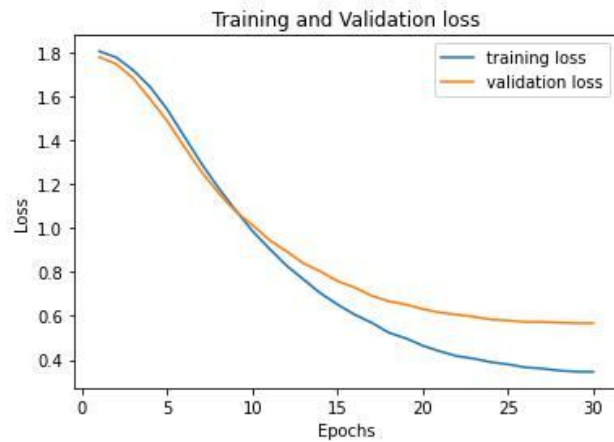


Fig 6. Training and Validation Loss

B. Analysis

Normalized confusion matrix						
sadness	0.70	0.00	0.01	0.17	0.11	0.00
joy	0.03	0.79	0.14	0.03	0.00	0.01
love	0.06	0.07	0.77	0.07	0.01	0.01
anger	0.03	0.00	0.00	0.83	0.14	0.00
fear	0.06	0.01	0.00	0.04	0.80	0.09
surprise	0.00	0.03	0.01	0.00	0.03	0.93
	sadness	joy	love	anger	fear	surprise

Fig 7. Confusion Matrix

classification report				
	precision	recall	f1-score	support
0	0.70	0.80	0.75	61
1	0.79	0.87	0.83	63
2	0.77	0.82	0.79	66
3	0.83	0.72	0.77	80
4	0.80	0.73	0.76	77
5	0.93	0.89	0.91	73
accuracy			0.80	420
macro avg	0.80	0.81	0.80	420
weighted avg	0.81	0.80	0.80	420

Fig 8. Classification Report

VI. CONCLUSIONS

The field of mental health is constantly evolving with new algorithms and techniques being introduced to address the challenges posed by psychiatric disorders. However, there is still significant scope for improvement and further exploration of these disorders using various machine-learning techniques. The selection of appropriate features is a critical factor that can significantly impact the accuracy of mental health classification models.

Recent studies have demonstrated the usefulness of machine learning in diagnosing and comprehending psychiatric disorders. The incorporation of different sensor modalities from advanced devices has opened up new avenues for identifying patients' mood states and other crucial features that can enhance mental health treatment outcomes. As the field of mental health research continues to progress, machine learning will undoubtedly play an increasingly crucial role in developing new and effective treatment approaches.

ACKNOWLEDGMENT

Our team conducted a project on predicting mental health using social media data and prescription audits. This project provided us with a valuable learning experience, allowing us to apply our theoretical knowledge practically. We extend our sincere gratitude to our mentor, Prof. Dr. (Mrs.) Gresha Bhatia from the Computer Department, for

her invaluable guidance, supervision, and support throughout the project. Her advice and encouragement were instrumental in the successful completion of our project.

Furthermore, we would like to thank Dr. (Mrs.) Nupur Giri, the Head of the Department, and Dr. (Mrs.) J.M. Nair, the principal, for providing us with the necessary resources to complete this project. We are also grateful for the support of our family and friends, whose unwavering belief in us and our goals inspired us to succeed. Their assistance and contributions were crucial to the success of this initiative.

REFERENCES

- [1] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019, June). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* (pp. 4171-4186). Association for Computational Linguistics.
- [2] Biddle, R., Joshi, A., Liu, S., Paris, C., & Xu, G. (2020, April). Leveraging sentiment distributions to distinguish figurative from literal health reports on twitter. In *Proceedings of The Web Conference 2020* (pp. 1217-1227).
- [3] Li, Y., Sun, L., Li, L., Li, P., Li, J., & Yu, H. (2020). BERT-based Mental Health Prediction on Social Media. In *International Conference on Computational Science* (pp. 446-458). Springer, Cham.
- [4] R.C.Kessler, "Screening for serious mental illness in the general population," *Arch. Gen. Psychiatry*, vol. 60, no. 2, pp. 184–189, 2003.
- [5] C.Chang, "All-cause mortality among people with serious mental illness (SMI), substance use disorders, and depressive disorders in southeast London: a cohort study," 2010.
- [6] M. Ernst, J. L. Gowin, C. Gaillard, R. T. Philips, and C. Grillon, "Brain Sciences Sketching the Power of Machine Learning to Decrypt a Neural Systems Model of Behavior," pp. 1–17, 2019.
- [7] S. F. Behaviors, A. A. Augustine, S. Vazire, D. Ph, N. Holtzman, and S. Gaddis, "Manifestations of Personality in Online Social Networks" vol.14, no. 9, 2011.
- [8] Kramer J, Conijn B, Ooijselaar P, Riper H. Effectiveness of a web-based solution-focused brief chat treatment for depressed adolescents and young adults: randomized controlled trial. *J Med Internet Res*. 2014.
- [9] Wang, X., Chen, J., & Zhang, X. (2019, June). A multi-task learning framework for mental health diagnosis with limited labeled social media data. In *2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE)* (pp. 1251-1256). IEEE.
- [10] Alghamdi, M., Hossain, M. S., & Muhammad, G. (2021, June). Detection of Depression and Anxiety Disorders using Machine Learning Algorithms. In *2021 IEEE International Conference on Power, Electronics and IoT (PEI)* (pp. 150-154). IEEE.
- [11] R. Jegan, S. Mathuranjani, and P. Sherly, "Mental Stress Detection and Classification using SVM Classifier: A Pilot Study," *2022 6th International Conference on Devices, Circuits and Systems (ICDCS)*, pp. 139-143, 2022, doi: 10.1109/ICDCS54290.2022.9780795.
- [12] T. Thamaraimealan et al., "Machine Learning based Patient Mental Health Prediction using Spectral Clustering and RBFN Algorithms," *2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS)*, 2022, pp. 1840-1843, doi: 10.1109/ICACCS54159.2022.9785142
- [13] Nguyen, T., Nguyen, T., Nguyen, H., Nguyen, V., & Dang, T. (2021, November). A Hybrid Approach for Identifying Mental Health Status from Social Media Data. In *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (pp. 3375-3380). IEEE.
- [14] S. Bhatia, A. Saha, S. Arora, and P. Bhattacharya, "Predicting Mental Health Status from Social Media Data using Deep Learning," *IEEE Access*, vol. 8, pp. 211578-211591, 2020.
- [15] R. Cao, S. Luo, S. Chen, D. Huang, and X. Zeng, "Predicting Mental Health from Social Media with Deep Learning," *IEEE Transactions on Multimedia*, vol. 21, no. 9, pp. 2310-2319, Sep. 2019.
- [16] Jaesong Lee, Joong-Hwi Shin, and Jun-Seok Kim. 2017. "Interactive visualization and manipulation of attention-based neural machine translation,". In *EMNLP: System Demonstrations*.
- [17] R. Pakravan, A. Nikfarjam, and G. R. Naiini, "Prediction of Mental Health Disorders Based on Social Media Texts and Network Analysis," *IEEE Transactions on Computational Social Systems*, vol. 6, no. 6, pp. 1397-1409, Dec. 2019.
- [18] J. M. Ramos-Rodríguez, M. A. Pérez-Rodríguez, J. A. Martínez-González, and J. F. Ortega-Rojas, "Detection of Depression in Twitter Users Based on the Frequency and Co-Occurrence of Medical Terms," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 3, pp. 899-906, May 2018.
- [19] N. Kaur and P. Khanna, "Detecting Depression and Anxiety Status of Users from Twitter Data using Deep Learning and Transfer Learning," *IEEE Access*, vol. 8, pp. 52056-52068, 2020.
- [20] S. Siddiqui, S. A. Hassan, A. Basit, M. A. Qureshi, and H. A. Baba, "Predicting Mental Health Conditions with Social Media Analysis," in *Proceedings of the IEEE International Conference on Big Data*, 2017, pp. 2777-2785.